(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 November 2001 (08.11.2001)

PCT

(10) International Publication Number
WO 01/83529 A2

(51) International Patent Classification[7]: C07K 14/095

(21) International Application Number: PCT/US01/13932

(22) International Filing Date: 28 April 2001 (28.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/200,298 28 April 2000 (28.04.2000) US

(71) Applicant *(for all designated States except US)*: **PLANET BIOTECHNOLOGY, INC.** [US/US]; 25571 Clawiter Road, Hayward, CA 94545 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LARRICK, James, William** [US/US]; Star Route Box #48, Woodside, CA 94062 (US). **WYCOFF, Keith, Lynn** [US/US]; 2399 Carmel Drive, Palo Alto, CA 94303 (US).

(74) Agents: **GUISE, Jeffrey, W.** et al.; Brobeck, Phleger & Harrison LLP, 12390 El Camino Real, San Diego, CA 92130 (US).

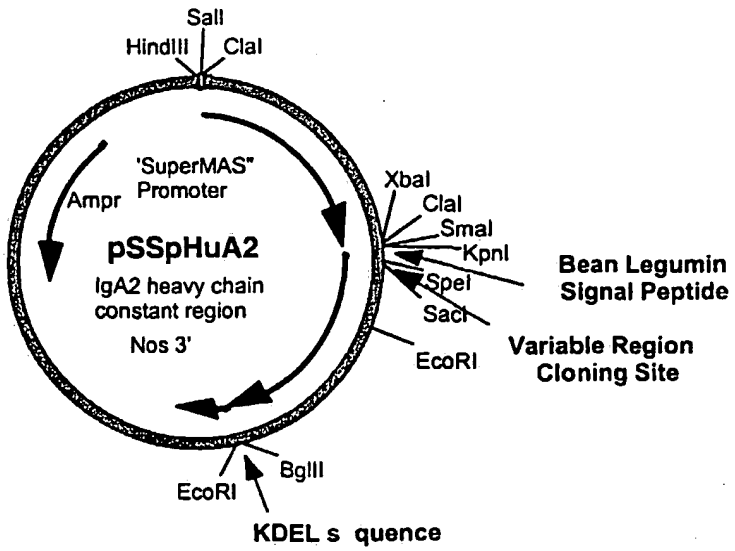
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL IMMUNOADHESIN FOR THE PREVENTION OF RHINOVIRUS INFECTION



(57) Abstract: The immunoadhesions of the present invention are useful in treating rhinovirus infections. The immunoadhesions contain a chimeric ICAM molecule and may optionally also contain J chain and secretory compounds. The chimeric ICAM molecule is a fusion protein that has a rhinovirus receptor protein linked to an immunoglobulin protein. This invention also includes the greatly increased and improved method of producing immunoadhesions in plants. Each of the components of an immunoadhesin is produced in a plant cell and thereby assembles within the plant cell. This method of producing the immunoadhesions of the present invention results in the efficient and economic production of these molecules. The present invention also contemplates the production of immunoadhesions in a variety of eukaryotic cells including plants and mammalian cells. The immunoadhesions of the present invention are useful as a therapeutic against the common cold in humans which is caused by rhinoviruses.

WO 01/83529 A2

NOVEL IMMUNOADHESIN FOR THE PREVENTION OF RHINOVIRUS INFECTION

PRIORITY

This application claims benefit under §119(e) of U.S. ***Provisional*** Patent Application No. 60/200,298, filed April 28, 2000, entitled NOVEL IMMUNOADHESIN FOR THE PREVENTION OF RHINOVIRUS INFECTION, and naming J. W. Larrick and K. L. Wycoff as inventors. This application is incorporated herein by reference in its entirety and for all purposes.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

Federal research support was provided in the form of an SBIR Phase I grant (R43 AI43122) and SBIR Phase II grant (2R44AI43122-02).

FIELD OF THE INVENTION

The present invention relates to immunoadhesins, fusions of the human rhinovirus receptor protein and immunoglobulin, and the expression of immunoadhesins in plants. The therapeutic use of immunoadhesins for the prevention and treatment of human rhinovirus infection is also contemplated.

BACKGROUND TO THE INVENTION

The common cold is generally a relatively mild disease, however significant complications resulting from colds, such as otitis media, sinusitis and asthma exacerbations are common. Human rhinoviruses (HRV) cause up to 50% of all adult colds and 25% of colds in children (Bella and Rossmann, *J Struct Biol.* 128:69-74, 1999, and Sperber and Hayden, *Antimicrob Agents Chemother.* 32:409-19, 1988). The cost to society runs into billions of dollar per year. These small, nonenveloped RNA viruses represent a subgroup of

picornavirus (Rueckert, *Virology*, pp. 507-548, eds. Fields, *et al.*, Raven Press, Ltd. New York, 1990) X-ray crystallography of rhinovirus identified a capsid 300 Å in diameter (1 Å = 0.1 nm) with icosahedral symmetry, constructed from sixty copies each of the viral coat proteins VP1, VP2, and VP3 (Rossmann, *Nature* 317:145-153, 1985). A surface depression or "canyon" on HRV was suggested as the receptor binding site (Colonno, *et al., Proc Natl Acad Sci U S A.* 85:5449-5453, 1985; Rossmann, *et al. Nature* 317:145-153, 1985). Of the 102 characterized HRV serotypes, 91 (known as the major group) share as their receptor a cell surface glycoprotein known as intercellular adhesion molecule-1 (ICAM-1) (Greve, *et al., Cell* 56:839-847, 1989; Staunton, *et al., Cell* 56:849-853, 1989); the binding site is located within N-terminal domain 1 (Greve, *et al., J Virol.* 65:6015-6023, 1991; Staunton, *et al., Cell* 61:243-254, 1990).

ICAM-1 is a membrane protein with five extracellular domains, a hydrophobic transmembrane domain, and a short cytoplasmic domain. ICAM-1 is expressed on many cells important in immune and inflammatory responses, and is inducible on others (Casasnovas, *et al., Proc Natl Acad Sci U S A.* 95:4134-9, 1998). ICAM-1 functions as a ligand for the leukocyte integrins LFA-1 and Mac-1 (Springer, *Cell.* 76:301-14, 1994; Staunton *et al., Cell* 61:243-254, 1990). On the cell surface, ICAM-1 is primarily a dimer due to association of the transmembrane domains (Miller, *et al., J Exp Med.* 182:1231-41, 1995; Reilly, *et al J Immunol.* 155:529-32, 1995).

Recombinant, soluble forms of ICAM-1 (sICAM-1) consisting of the five extracellular domains were shown to be effective in blocking rhinovirus infection of human cells *in vitro* (Greve, *et al., J Virol.* 65:6015-6023, 1991; Marlin, *et al., Nature.* 344:70-2, 1990). Evaluation of sICAM-1 activity against a spectrum of laboratory strains and field isolates showed that all major strains of HRV are sensitive to sICAM-1. Minor strains, which do not use ICAM as a receptor, were unaffected by sICAM-1 (Crump *et al., Antiviral Chem Chemother.* 4:323-327, 1993; Ohlin, *et al., Antimicrob Agents Chemother.* 38:1413-5, 1994).

The anti-viral activity of soluble ICAM-1 *in vitro* appears to be mediated by more than one mechanism. These mechanisms include competition with cell-surface ICAM-1 for binding sites, interference with virus entry or uncoating, and direct inactivation by premature release of viral RNA and formation of empty capsids (Arruda, *et al., Antimicrob Agents Chemother.* 36:1186-1191, 1992; Greve, *et al., J Virol.* 65:6015-6023, 1991; Marlin, *et al., Nature* 344:70-2, 1990; Martin *et al., J Virol.* 67:3561-8, 1993).

The host range of HRV is restricted to primates. A recent study showed that soluble ICAM-1 was effective in preventing rhinovirus infection in chimpanzees (Huguenel, *et al., Am J Respir Crit Care Med.* 155:1206-10, 1997). Although chimpanzees do not show clinical symptoms, infection was demonstrated by measuring seroconversion and virus shedding. A single dose of 10 mg of soluble ICAM-1 as an intranasal spray was effective at preventing infection by HRV-16 when co-administered with HRV, or when the virus was administered ten minutes later.

A human clinical trial with soluble ICAM-1 showed that it reduced the severity of experimental HRV colds (Turner, *et al., JAMA* 281:1797-804, 1999). In this trial a total of 196 subjects received either soluble ICAM-1 or placebo in various formulations. Some subjects were given soluble ICAM-1 or placebo starting seven hours before inoculation with HRV 39 and others were started twelve hours after virus inoculation. Medications were administered as either an intranasal solution or powder, given in six daily doses for seven days (a total of 4.4 mg per day). In this study, soluble ICAM-1 did not prevent infection, as measured by either virus isolation or seroconversion (infection rate of 92% for placebo-treated vs. 85% of soluble ICAM-1 treated). However, soluble ICAM-1 did have an impact on all measures of illness. The total symptom score was reduced by 45%, the proportion of subjects with clinical colds was reduced 23% and nasal mucus weight was reduced by 56%. There was not a significant difference between the use of powder or solution formulations, or between pre- and post-inoculation groups. Treatment with soluble ICAM-1 did not result in any adverse effects or evidence of absorption through the nasal

mucosa. Also, there was no inhibition of the development of anti-HRV type-specific antibodies.

As discussed, ICAM-1 is dimeric on the cell surface. Martin *et al.*, in *J Virol.* 67:3561-8, (1993) first proposed that multivalent binding to HRV by a multimeric soluble ICAM might result in a higher effective affinity, termed avidity, and thus facilitate uncoating of the virus. They constructed multivalent, ICAM-1/immunoglobulin molecules, postulating that these would be more effective than monovalent soluble ICAM-1 in neutralizing HRV and thus would have increased therapeutic utility. These ICAM-1/immunoglobulin molecules included ICAM-1 amino-terminal domains 1 and 2 fused to the hinge and constant domains of the heavy chains of IgA1 (IC1-2D/IgA), IgM (IC1-2D/IgM) and IgG1(IC1-2D/IgG). In addition, five extracellular domains were fused to IgA1 (IC1-5D/IgA). These ICAM-1/immunoglobulin molecules were compared with soluble forms of ICAM-1 having two (sIC1-2D) and five (sIC1-5D) domains in assays of HRV binding, infectivity and conformation. The ICAM-1/IgA immunoglobulin (IC1-5D/IgA) was 200 times, and the ICAM-1/IgM immunoglobulin (IC1-2D/IgM) and ICAM-1/IgG immunoglobulin molecules (IC1-2D/IgG) were 25 and 10 times, more effective than soluble ICAM-1. These molecules were highly effective in inhibiting rhinovirus binding to cells and disrupting the conformation of the virus capsid. The ICAM-1/IgA immunoglobulin molecules were effective in the nanomolar concentration range. Comparison of IC1-2D/IgA and IC1-2D/IgG showed that the class of Ig constant region used had a large impact on efficacy.

A subsequent study compared the inhibitory activities of soluble ICAM-1 and IC1-5D/IgA against nine major HRV serotypes and a variant of HRV-39 selected for moderate resistance to soluble ICAM-1 (Crump, *et al., Antimicrob Agents Chemother.* 38:1425-7, 1993). IC1-5D/IgA was more potent than monomeric soluble ICAM-1 by 50 to 143 times on a weight basis and by 60 to 170 times on a molar basis against the standard serotypes. The HRV-39 variant was 38-fold more resistant to soluble ICAM-1 than the wild-type, and it was only 5-fold more resistant to IC1-5D/IgA. This is consistent with the hypothesis that virus escape from inhibition by multivalent molecules would be expected to

occur at lower frequency than virus escape from inhibition by monomeric soluble receptor (Martin, *et al., J Virol.* 67:3561-8, 1993). An assay designed to measure viral inactivation showed that HRV-39 and HRV-13 were not directly inactivated to a significant extent by soluble ICAM-1 (<0.5 $\log_{10}$ reduction in infectivity). However, incubation with IC1-5D/IgA resulted in a reduction of infectivity of these same viruses by about 1.0 $\log_{10}$ (Crump, *et al., Antimicrob Agents Chemother.* 38:1425-7, 1994). Results by Martin *et al.* (*J Virol.* 67:3561-8, 1993) suggest that the greater the valence, the greater the effectiveness of the molecules. Dimeric and decameric forms of IC1-2/IgM were separable by sucrose gradient sedimentation. The decameric form was five times more effective than the dimeric form at blocking binding of HRV to HeLa cells.

The ICAM-1/immunoglobulin molecules that have been described suffer from several drawbacks, including the laborious production techniques and high costs associated with those production methods. In addition, the previously described ICAM-1/immunoglobulin molecules have limited stability as multimers in the harsh environment in which the molecule must inactivate rhinoviruses.

The immunoadhesins of the present invention have significant advantages over what has been described in the art. The immunoadhesins of the present invention that are expressed in plants would be tetrameric, rather than only dimeric. Immunoadhesins having multiple binding sites have a higher effective affinity for the virus, thereby increasing the effectiveness of the immunoadhesin. In addition, the association of secretory component and immunoglobulin J chain with the immunoadhesin of the present invention increases the stability of the immunoadhesin in the mucosal environment (Corthesy, *Biochem Soc Trans.* 25:471-475, 1997). Secretory IgA, which is associated with secretory component, is the antibody isotype normally found in mucosal secretions, including milk and colostrum. Unlike other antibody isotypes, SIgA can pass through the gut with very little proteolytic degradation. It also is very stable in crude plant preparations at room temperature. A function of the secretory component appears to be to protect the antibody from the harsh environment of the mucosa (Paul, *Fundamental Immunology*, Raven Press, NY, Third

Edition, pp. 303-304, 1993). Furthermore, the immunoadhesin of the present invention are significantly less expensive to produce in plants than in animal cell culture, and production in plants would make it safer for human use, since plants are not known to harbor any animal viruses.

SUMMARY OF THE INVENTION

The present invention contemplates an immunoadhesin comprising a chimeric ICAM-1 molecule having a rhinovirus receptor protein linked to at least a portion of an immunoglobulin heavy chain, wherein J chain and secretory component are associated with the chimeric ICAM-1 molecule.

In preferred embodiments, the immunoadhesin of the present invention is comprised of a rhinovirus receptor protein made of any combination of extracellular domains 1, 2, 3, 4 and 5 of the rhinovirus receptor protein, ICAM-1, linked to an immunoglobulin heavy chain. Also contemplated by the present invention are immunoadhesins of the present invention in which the immunoglobulin is IgA, IgA_1, IgA_2, IgG_1, IgG_2, IgG_3, IgG_4, IgM, IgD, IgE or a chimeric immunoglobulin heavy chain made up of domains or segments from different immunoglobulin isotypes.

In other preferred embodiments of the present invention, the immunoadhesin comprises multiple chimeric ICAM-1 molecules associated with J chain and secretory component. The increase in valency results in a higher effective affinity for the rhinovirus, thereby increasing the effectiveness of the immunoadhesin.

In a preferred embodiment of the present invention, all proteins used to make the immunoadhesin of the present invention are human proteins. In addition to production in plants or plant cells, the present invention contemplates an immunoadhesin expressed in mammalian cells, hairy root cultures, plant cells in tissue culture, and heterologous cells derived from plants, vertebrates or invertebrates.

In preferred embodiments of the present invention, the immunoadhesins are expressed, in plants, including monocotyledonous plants and dicotyledonous plants as a part of the plants genome. Expression in plants, as opposed to expression in cultured cells, allows for a significant reduction in the cost of producing the immunoadhesin.

The present invention contemplates an immunoadhesin having plant-specific glycosylation. A gene coding for a polypeptide having within its amino acid sequence, the glycosylation signal asparagine-X-serine/threonine, where X can be any amino acid residue, is glycosylated via oligosaccharides linked to the asparagine residue of the sequence when expressed in a plant cell. See Marshall, *Ann. Rev. Biochem.*, 41:673 (1972) and Marshall, *Biochem. Soc. Symp.*, 40:17 (1974) for a general review of the polypeptide sequences that function as glycosylation signals. These signals are recognized in both mammalian and in plant cells. At the end of their maturation, proteins expressed in plants or plant cells have a different pattern of glycosylation than do proteins expressed in other types of cells, including mammalian cells and insect cells. Detailed studies characterizing plant-specific glycosylation and comparing it with glycosylation in other cell types have been performed, for example, in studies described by Cabanes-Macheteau *et al.*, *Glycobiology* 9(4):365-372 (1999), and Altmann, *Glycoconjugate J.* 14:643-646 (1997). These groups and others have shown that plant-specific glycosylation generates glycans that have xylose linked β(1,2) to mannose, but xylose is not linked β(1,2) to mannose as a result of glycosylation in mammalian and insect cells. Plant-specific glycosylation results in a fucose linked α(1,3) to the proximal GlcNAc, while glycosylation in mammalian cells results in a fucoselinked α(1,6) to the proximal GlcNAc. Furthermore, plant-specific glycosylation does not result in the addition of a sialic acid to the terminus of the protein glycan, whereas in glycosylation in mammalian cells, sialic acid is added.

In other embodiments, the immunoadhesin of the present invention is part of a composition comprising plant material and the immunoadhesin, associated with J chain and secretory component. The plant material present may be plant cell walls, plant organelles, plant cytoplasms, intact plant cells, viable plants, and the like. The particular plant materials

or plant macromolecules that may be present include ribulose bisphosphate carboxylase, light harvesting complex, pigments, secondary metabolites or chlorophyll. Compositions of the present invention may have an immunoadhesin concentration of between 0.001% and 99.9% mass excluding water. In other embodiments, the immunoadhesin is present in a concentration of 0.01% to 99% mass excluding water. In other embodiments, the compositions of the present invention have plant material or plant macromolecules present at a concentration of 0.01% to 99% mass excluding water.

The present invention also contemplates methods for the treatment or prevention of human rhinovirus infection in a subject, including reducing the infection by human rhinovirus of host cells susceptible to infection by the virus, or reducing the initiation or spread of the common cold due to human rhinovirus, by a method comprising contacting the virus with an immunoadhesin of the present invention, wherein the immunoadhesin binds to the human rhinovirus and reduces infectivity. The immunoadhesin could mediate infection by competition with cell-surface ICAM-1 for binding sites, interference with virus entry or uncoating, and/or direct inactivation by premature release of viral RNA and formation of empty capsids (Arruda, *et al., Antimicrob. Agents Chemother.* 36:1186-1191, 1992; Greve, *et al., J. Virol.* 65:6015-6023, 1991; Martin, *et al., Nature* 344:70-2, 1990; Martin, *et al., J. Virol.* 67:3561-8, 1993). In another embodiment, human rhinovirus infection in a subject is treated by a method comprising intranasally administering to the subject an effective amount of an immunoadhesin of the present invention, wherein the immunoadhesin reduces human rhinovirus infectivity.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates pSSPHuA2, vector in which DNAs encoding a chimeric ICAM-1 molecule containing the first five domains of human ICAM-1 and the Fc region of human IgA2m(2) were fused. This vector contains the SuperMas promoter for driving the expression of a signal peptide and the constant regions of the human IgA2m(2) heavy-chain. Sequences encoding ICAM domains 1-5 were amplified, by PCR, to contain convenient restriction sites

(5' SpeI and 3' Spe I) for insertion between the signal peptide and the Cα1 domain. DNA encoding an ER retention signal (RSEKDEL) was appended to the 3' end of the heavy-chain in order to boost the expression level of the construct.

FIG. 2 illustrates a chimeric ICAM molecule. **2A** shows the DNA expression cassette from which the chimeric ICAM-1 molecule was produced. **2B** shows the amino acid sequence, after signal peptide cleavange, of the mature form of the fusion protein. Amino acids introduced by the cloning procedure are underlined and mark the junction between the five extracellular domains of ICAM-1 and the Cα1-Cα3 domains of the IgA2m(2) heavy chain. The bolded N's indicate the fifteen potential glycosylation sites.

FIG. 3 illustrates the expression of the immunoadhesin in independently transformed tobacco calli. **3A** shows immunoblots of non-reducing SDS-polyacrylamide gels on which samples containing different transformed tobacco calli (**C**) and aqueous extracts (**Aq**) were run and probed for the presence of human ICAM. The molecular weight markers are indicated, and the reference standard (**R**) was a mixture (~75 ng each) of human ICAM (~75 kD) and human SigA (>>250 kD). **3B** shows immunoblots of nonreducing SDS-polyacrylamide gels containing various fractions of partially purified immunoadhesin from callus Rhi107-11. The purification fractions analyzed were juice (**J**), G-100 fraction (**G**), sterile filtered G-100 fraction (**SG**), and a mixture of reference standards of human SigA (75 ng) and human ICAM-1 (75ng) (**RS**).

Blots were probed with antibodies against human ICAM (**~ICAM**), human IgA heavy chain (**~α**), human secretory component (**~SC**) and human J chain (**~J**). Secondary, enzyme-conjugated antibodies were employed as necessary to label immuno-positive bands with alkaline phosphatase.

FIG. 4 illustrates the results of an enzyme-linked immunosorbent assay (ELISA) showing competition between plant extract and soluble ICAM-1 for binding to an anti-ICAM mAb. For the assay, 96-well plates were coated with 0.25 μg soluble ICAM-1/ml. The squares represent the increasing concentrations of sICAM and the circles represent the

increasing amounts of callus extract (sterile filtered fraction from G-100) used to compete with the adhered ICAM for a constant amount of a mouse (anti-human ICAM) antibody.

FIG. 5 illustrates the results of an assay showing the ability of an immunoadhesin to inhibit human rhinovirus killing of HeLa cells (cytopathic effect, or CPE, assay). **5A** shows the results of an assay comparing the CPE of human rhinovirus on HeLa cells in the presence of partially purified extracts containing either the immunoadhesin in the ICAM-Fc fusion (IC1-5D/IgA) or containing an antibody against doxorubicin. (The right side-up and upside-down triangles represent two extracts derived from Rhi107-11, containing the immunoadhesin.) **5B** shows the results of an assay comparing the CPE of human rhinovirus on HeLa cells in the presence of soluble human ICAM-1 or an extract from the immunoadhesin in the ICAM-Fc fusion (IC1-5D/IgA). The Inset shows scale expansion in the range of the IC50 for soluble ICAM (1.35 μg/ml) and for IC1-5D/IgA (0.12 μg/ml; 11.3 fold-less).

FIG. 6 shows an evaluation of the production necessities for making 1 gram of finished immunoadhesin. In this diagram, the number of plants needed for 1 g of immunoadhesin, at 20% yield, at expected levels of expression and plant weight is illustrated. At different levels of immunoadhesin expression (mg/kg fresh weight) and overall recovery (set at 20%), the weight of each plant, and so the total number of plants, may be determined for a specified production target (1 g/harvest) within a window (dotted square) of reasonable possibilities. The number of required plants decreases, inversely, with the number of specified growth and re-growth periods. The expected biomass production, a function of time and growth conditions, influences the time to harvest and the time between harvests. These growth periods can be adjusted to the realities of the purification schedule by staggering planting and harvesting dates.

FIG. 7 shows the coding and amino acid sequences of each of the immunoglobulin genes and proteins listed in Table 2.

FIG. 8 shows the sequences of plasmids used to transform plants, as described in Example 2, for use in studies of the expression of immunoadhesins of the present invention.

FIG. 8 A shows the nucleotide and protein sequences for plasmids PSSpICAMHuA2

FIG 8 B shows the nucleotide protein sequences for the bean legumin signal peptide.

FIG 8 C shows the nucleotide and amino acid sequence of the protein coding region of pSHuJ.

FIG 8 D shows the nucleotide and amino acid sequence of protein coding region of pSHuSC.

FIG 8 E shows the nucleotide sequence of plasmids pBMSP-1.

FIG 8 F shows the nucleotide sequence of plasmids pBMSP-1spJSC.

FIG. 9 contains nucleotide and protein sequences SEQ ID NO: 1; SEQ ID NO: 2; SEQ ID NO: 3; SEQ ID NO: 4; SEQ ID NO: 5; SEQ ID NO: 6; SEQ ID NO: 7; SEQ ID NO: 8, for ICAM-1, and human IgA_2 and other nucleotide sequences.

DETAILED DESCRIPTION OF THE INVENTION

A. Definitions.

As used herein, the following abbreviations and terms include, but are not necessarily limited to, the following definitions.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of immunology, molecular biology, microbiology, cell biology and recombinant DNA, which are within the skill of the art. See, e.g., Sambrook, *et al., Molecular Cloning: A Laboratory Manual, 2nd edition* (1989); *Current Protocols In Molecular Biology (F.M. Ausubel, et al. eds.*, (1987)); the series *Methods In Enzymology* (Academic Press, Inc.); M.J. MacPherson, *et al.*, eds. *Pcr 2: A Practical Approach* (1995);

Harlow and Lane, eds, *Antibodies: A Laboratory Manual* (1988), and H. Jones, *Methods In Molecular Biology vol. 49, "Plant Gene Transfer And Expression Protocols"* (1995).

Immunoglobulin molecule or Antibody. A polypeptide or multimeric protein containing the immunologically active portions of an immunoglobulin heavy chain and immunoglobulin light chain covalently coupled together and capable of specifically combining with antigen. The immunoglobulins or antibody molecules are a large family of molecules that include several types of molecules such as IgD, IgG, IgA, secretory IgA (SIgA), IgM, and IgE.

Construct or Vector. An artificially assembled DNA segment to be transferred into a target plant tissue or cell. Typically, the construct will include the gene or genes of a particular interest, a marker gene and appropriate control sequences. The term "plasmid" refers to an autonomous, self-replicating extrachromosomal DNA molecule. In a preferred embodiment, the plasmid constructs of the present invention contain sequences coding for heavy and light chains of an antibody. Plasmid constructs containing suitable regulatory elements are also referred to as "expression cassettes." In a preferred embodiment, a plasmid construct can also contain a screening or selectable marker, for example an antibiotic resistance gene.

Selectable marker. A gene that encodes a product that allows the growth of transgenic tissue on a selective medium. Non-limiting examples of selectable markers include genes encoding for antibiotic resistance, e.g., ampicillin, kanamycin, or the like. Other selectable markers will be known to those of skill in the art.

Transgenic plant. Genetically engineered plant or progeny of genetically engineered plants. The transgenic plant usually contains material from at least one unrelated organism, such as a virus, another plant or animal.

Chimeric ICAM-1 molecule: The fusion of any combination of the extracellular domains 1, 2, 3, 4 and 5 of ICAM-1 with at least a part of an immunoglobulin heavy chain

protein, made by linking ICAM-1 sequence upstream of an immunoglobulin heavy chain gene sequence and expressing the encoded protein from the construct.

Chimeric immunoglobulin heavy chain: An immunoglobulin derived heavy chain having at least a portion of its amino acid sequence derived from an immunoglobulin heavy chain of a different isotype or subtype or some other peptide, polypeptide or protein. Typically, a chimeric immunoglobulin heavy chain has its amino acid residue sequence derived from at least two different isotypes or subtypes of immunoglobulin heavy chain.

Dicotyledonous plants (dicots): Flowering plants whose embryos have two seed halves or cotyledons. Examples of dicots are: tobacco; tomato; the legumes including alfalfa; oaks; maples; roses; mints; squashes; daisies; walnuts; cacti; violets and buttercups.

Effective amount: An effective amount of an immunoadhesin of the present invention is sufficient to detectably inhibit rhinovirus infection, cytotoxicity or replication; or to reduce the severity or length of rhinovirus infection.

Human rhinovirus (HRV): A nonenveloped RNA virus representing a subgroup of picornavirus, that is a major cause of the common cold in humans. Rhinoviruses are described in Rhinoviruses, Reoviruses, and Parvoviruses, pp. 1057-1059, *Zinsser Microbiology*, Joklik *et al.*, eds. Appleton and Lange (1992).

Immunoadhesin : A complex containing a chimeric ICAM-1 molecule, and optionally containing secretory component, and J chain.

Immunoglobulin heavy chain: A polypeptide that contains at least a portion of the antigen binding domain of an immunoglobulin and at least a portion of a variable region of an immunoglobulin heavy chain or at least a portion of a constant region of an immunoglobulin heavy chain. Thus, the immunoglobulin derived heavy chain has significant regions of amino acid sequence homology with a member of the immunoglobulin gene superfamily. For example, the heavy chain in an Fab fragment is an immunoglobulin-derived heavy chain.

Immunoglobulin light chain: A polypeptide that contains at least a portion of the antigen binding domain of an immunoglobulin and at least a portion of the variable region or at least a portion of a constant region of an immunoglobulin light chain. Thus, the immunoglobulin-derived light chain has significant regions of amino acid homology with a member of the immunoglobulin gene superfamily.

Immunoglobulin molecule: A protein containing the immunologically-active portions of an immunoglobulin heavy chain and immunoglobulin light chain covalently coupled together and capable of specifically combining with antigen.

ICAM-1: Intercellular adhesion molecule-1. In humans, ICAM-1 functions as the receptor for human rhinovirus.

J chain: A polypeptide that is involved in the polymerization of immunoglobulins and transport of polymerized immunoglobulins through epithelial cells. See, *The Immunoglobulin Helper: The J Chain in Immunoglobulin Genes*, at pg. 345, Academic Press (1989). J chain is found in pentameric IgM and dimeric IgA and typically attached via disulphide bonds. J chain has been studied in both mouse and human.

Monocotyledonous plants (monocots): Flowering plants whose embryos have one cotyledon or seed leaf. Examples of monocots are: lilies; grasses; corn; grains, including oats, wheat and barley; orchids; irises; onions and palms.

Glycosylation: The modification of a protein by oligosaccharides. See, Marshall, *Ann. Rev. Biochem.*, 41:673 (1972) and Marshall, *Biochem. Soc. Symp.*, 40:17 (1974) for a general review of the polypeptide sequences that function as glycosylation signals. These signals are recognized in both mammalian and in plant cells.

Plant-specific glycosylation: The glycosylation pattern found on plant-expressed proteins, which is different from that found in proteins made in mammalian or insect cells. Proteins expressed in plants or plant cells have a different pattern of glycosylation than do proteins expressed in other types of cells, including mammalian cells and insect cells.

Detailed studies characterizing plant-specific glycosylation and comparing it with glycosylation in other cell types have been performed by Cabanes-Macheteau *et al.*, *Glycobiology* 9(4):365-372 (1999), Lerouge *et al.*, *Plant Molecular Biology* 38:31-48 (1998) and Altmann, *Glycoconjugate J.* 14:643-646 (1997). Plant-specific glycosylation generates glycans that have xylose linked β(1,2) to mannose. Neither mammalian nor insect glycosylation generate xylose linked β(1,2) to mannose. Plants do not have a sialic acid linked to the terminus of the glycan, whereas mammalian cells do. In addition, plant-specific glycosylation results in a fucose linked α(1,3) to the proximal GlcNAc, while glycosylation in mammalian cells results in a fucose linked α(1,6) to the proximal GlcNAc.

Secretory component (SC): A component of secretory immunoglobulins that helps to protect the immunoglobulin against inactivating agents thereby increasing the biological effectiveness of secretory immunoglobulin. The secretory component may be from any mammal or rodent including mouse or human.

sICAM: A naturally-occurring soluble truncated form of ICAM-1 lacking both the hydrophobic transmembrane domain and the carboxy-terminal cytoplasmic domain of ICAM.

The articles, patents and patent applications cited in this document are incorporated into this document as if set forth in full.

B. Immunoadhesins Containing Chimeric ICAM Molecules.

The present invention provides novel methods for producing immunoadhesin molecules containing chimeric ICAM molecules. The immunoadhesins of the present invention contain chimeric ICAM-1 molecules made up of a rhinovirus receptor protein linked to a portion of an immunoglobulin heavy chain molecule in association with J chain and secretory component. The chimeric ICAM-1 molecules of the present invention contain two molecules derived from different sources: a rhinovirus receptor protein portion and an immunoglobulin chain portion. The rhinovirus receptor protein of the present invention is derived from the intercellular adhesion molecule 1 (ICAM-1). The nucleotide sequence for

the human rhinovirus receptor ICAM-1 has been determined and characterized by Staunton, *et al.*, *Cell* 52:925-933 (1988); Greve, *et al. Cell* 56:839-847 (1989); Greve, *et al. J. Virology* 65:6015-6023 (1991); Staunton, *et al.*, *Cell*, 61:243-254 (1990) and described in Sequence ID No. 3 and GenBank accession no. M24283.

The ICAM-1 molecule is a membrane protein (SEQ ID NOS: 1 and 2) that has 5 extracellular domains, a hydrophobic transmembrane domain and a short cytoplasmic domain. These features have been described by Casasnovas, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 95:4134-4139 (1998) and Staunton, *et al*, Cell 52:925-933 (1988). Of particular use in the present invention are the domains of the ICAM-1 molecule that are responsible for the binding of human rhinoviruses which have been localized to the N-terminal domains 1 and 2 (Greve, *et al.*, *J. Virol.*, 65:6015-6023 1991, and Staunton, *et al.*, *Cell*, 61:243-245 1990. The present invention also contemplates rhinovirus receptor protein portions which include any combination of extracellular domains 1, 2, 3, 4, and 5 of the ICAM-1 molecule. In particular preferred embodiments, the rhinovirus receptor protein portion includes domains 1 and 2 of the ICAM-1 molecule and in other preferred embodiments domains 1, 2, 3, 4 and 5 of the ICAM-1 molecule are present.

The boundaries of the 5 extracellular domains are well known in the art and described in Staunton, *et al.*, Cell 52:925-933 (1988). The approximated domain boundaries are shown in Table 1 below.

Table 1

ICAM-1 Domains	Amino Acids
1	1-88
2	89-105
3	106-284
4	285-385
5	386-453

As used in the present invention, the ICAM-1 domain 1 is from about residue 1 to about residue 88; domain 2 is from about residue 89 to about residue 105; domain 3 is from about residue 106 to about residue 284; domain 4 is from about residue 285 to about 385; and domain 5 is from about residue 386 to 453. One of skill in the art will understand that the exact boundaries of these domains may vary.

The chimeric ICAM-1 molecules of the present invention preferably contain at least a portion of an IgM or IgA heavy chain which allows that immunoglobulin heavy chain to bind to immunoglobulin J chain and thereby binds to the secretory component. It is contemplated that the portion of the chimeric ICAM-1 molecule derived from the immunoglobulin heavy chain of the present invention may be comprised of individual domains selected from the IgA heavy chain or the IgM heavy chain or from some other isotype of heavy chain. It is also contemplated that an immunoglobulin domain derived from an immunoglobulin heavy chain other than IgA or IgM or from an immunoglobulin light chain may be molecularly engineered to bind immunoglobulin J chain and thus may be used to produce immunoglobulins of the present invention.

One skilled in the art will understand that immunoglobulins consist of domains which are approximately 100-110 amino acid residues. These various domains are well known in the art and have known boundaries. The removal of a single domain and its replacement with a domain of another antibody molecule is easily achieved with modern molecular biology. The domains are globular structures which are stabilized by intrachain disulfide bonds. This confers a discrete shape and makes the domains a self-contained unit that can be replaced or interchanged with other similarly shaped domains. The heavy chain constant region domains of the immunoglobulins confer various properties known as antibody effector functions on a particular molecule containing that domain. Example effector functions include complement fixation, placental transfer, binding to staphyloccal protein, binding to streptococcal protein G, binding to mononuclear cells, neutrophils or mast cells and basophils. The association of particular domains and particular immunoglobulin isotypes with these effector functions is

well known and for example, described in Immunology, Roitt *et al.*, *Mosby St. Louis, Mo.* (1993 3rd Ed.)

One of skill in the art will be able to identify immunoglobulin heavy chain constant region sequences. For example, a number of immunoglobulin DNA and protein sequences are available through GenBank. Table 2 shows the GenBank Accession numbers of immunoglobulin heavy chain genes and the proteins encoded by the genes. The sequences listed in Table 2 are shown in Fig. 7.

Table 2

GENBANK ACCESSION NO.	HUMAN IMMUNOGLOBULIN SEQUENCE NAME	SEQ ID NO.
J00220	$Ig_{\alpha 1}$ Heavy Chain Constant Region Coding Sequence	15
J00220	$Ig_{\alpha 1}$ Heavy Chain Constant Region Amino Acid Sequence	16
J00221	IgA_2 Heavy Chain Constant Region Coding Sequence	17
J00221	IgA_2 Chain Constant Region Amino Acid Sequence	18
J00228	$Ig_{\gamma 1}$ Heavy Chain Constant Region Coding Sequence	19
J00228	$Ig_{\gamma 1}$ Heavy Chain Constant Region Amino Acid Sequence	20
J00230 V00554	IgG_2 Heavy Chain Constant Region Coding Sequence	21
J00230 V00554	IgG_2 Heavy Chain Constant Region Amino Acid Sequence	22
X03604 M12958	IgG_3 Heavy Chain Constant Region Coding Sequence	23
X03604 M12958	IgG_3 Heavy Chain Constant Region Amino Acid Sequence	24
K01316	IgG_4 Heavy Chain Constant Region Coding Sequence	25
K01316	IgG_4 Heavy Chain Constant Region Amino Acid Sequence	26
K02876	IgD Heavy Chain Constant Region Coding Sequence	27
K02876	IgD Heavy Chain Constant Region Acid Sequence	28
K02877	IgD Heavy Chain Constant Region Coding Sequence	29
K02877	IgD Heavy Chain Constant Region Amino Acid Sequence	30

GENBANK ACCESSION NO.	HUMAN IMMUNOGLOBULIN SEQUENCE NAME	SEQ ID NO.
K02878	Germline IgD Heavy Chain Coding Sequence	31
K02878	Germline IgD Heavy Chain Amino Acid Sequence	32
K02879	Germline IgD Heavy Chain C-δ-3 Domain Coding Sequence	33
K02879	Germline IgD Heavy Chain C-δ-3 Amino Acid Sequence	34
K01311	Germline IgD Heavy Chain J-δ Region: C-δ CH1 Amino Acid Sequence	35
K02880	Germline IgD Heavy Chain Gene, C-Region, Secreted Terminus Coding Sequence	36
K02880	Germline IgD Heavy Chain Gene, C-Region, Secreted Terminus Amino Acid Sequence	37
K02881	Germline IgD-Heavy Chain Gene, C-Region, First Domain of Membrane Terminus Coding Sequence	38
K02881	Germline IgD-Heavy Chain Gene, C-Region, First Domain of Membrane Terminus Amino Acid Sequence	39
K02882	Germline IgD Heavy Chain Coding Sequence	40
K02882	Germline IgD Heavy Chain Amino Acid Sequence	41
K02875	Germline IgD Heavy Chain Gene, C-Region, C-δ-1 Domain Coding Sequence	42
K02875	Germline IgD Heavy Chain Gene, C-Region, C-δ-1 Domain Amino Acid Sequence	43
L00022 J00227 V00555	IgE Heavy Chain Constant Region Coding Sequence	44
L00022 J00227 V00555	IgE Heavy Chain Constant Region Amino Acid Sequence	45
X17115	IgM Heavy Chain Complete Sequence Coding Sequence	46
X17115	IgM Heavy Chain Complete Sequence mRNA	47

The immunoadhesins of the present invention may, in addition to the chimeric ICAM-1 molecule, contain immunoglobulin light chains, or immunoglobulin J chain bound to the immunoglobulin derived heavy chains. In preferred embodiments, the immunoadhesin of the present invention comprises two or four chimeric ICAM-1 molecules and an immunoglobulin J chain bound to at least one of the chimeric ICAM-1 molecules. The J chain is described and known in the art. See, for example, M. Koshland, The Immunoglobulin Helper: The J Chain, in Immunoglobulin Genes, Academic Press, London, pg. 345, (1989) and Matsuuchi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 83:456-460 (1986). The sequence of the immunoglobulin J chain is available on various databases in the United States.

The immunoadhesin of the present invention may have a secretory component associated with the chimeric ICAM-1 molecule. This association may occur by hydrogen bonds, disulfide bonds, covalent bonds, ionic interactions or combinations of these various bonds. Typically, chimeric ICAM-1 molecules are held together by disulfide bonds between the molecules. The interaction of the chimeric ICAM-1 molecules may be non-covalent or disulfide bonding. The present invention contemplates the use of secretory component from a number of different species, including human, rat, rabbit, bovine and the like. The nucleotide sequences for these molecules are well known in the art. For example, U.S. Patent 6,046,037 contains many of the sequences and this patent is incorporated herein by reference.

The immunoadhesins of the present invention containing the secretory component, the chimeric ICAM-1 molecule and J chain are typically bonded together by one of the following: hydrogen bonds, disulfide bonds, covalent bonds, ionic interactions or combinations of these bonds.

The present invention also contemplates immunoadhesins which comprise more than one chimeric ICAM-1 molecule. The immunoadhesin may contain chimeric ICAM-1 molecules that are monomeric units and not disulfide bonded to other chimeric ICAM-1 molecules. In preferred embodiments, the immunoadhesion does contain chimeric ICAM-1

molecules that are in association with other chimeric ICAM-1 molecules to form dimers and other multivalent molecules. Typically the chimeric ICAM-1 molecule is present as a dimer because of the association of the immunoglobulin portion of the chimeric molecule. The immunoglobulin portion of the chimeric ICAM-1 molecule allows the association of two chimeric ICAM-1 molecules to form a dimeric molecule having two active binding portions made up of the rhinovirus receptor protein portion. In preferred embodiments, dimerization occurs via the disulfide bonding regions that normally occur between the immunoglobulin domains as part of a naturally-occurring immunoglobulin molecule and the native immunoglobulin protein. One of skill in the art will understand that these disulfide bonds that are normally present in the native immunoglobulin molecule can be modified, moved and removed while still maintaining the ability to form a dimer of the chimeric ICAM-1 molecules.

In other preferred embodiments, the immunoadhesin contains multimeric forms of the chimeric ICAM-1 molecule due to the association of J chain with the immunoglobulin portion of the chimeric ICAM molecule. The association of J chain with the dimer of two chimeric ICAM-1 molecules allows the formation of tetrameric forms of the immunoadhesin. In a preferred embodiment, the immunoglobulin portion of the chimeric ICAM-1 molecule is derived from the IgA molecule, and the addition of J chain allows the formation of a tetrameric complex containing four chimeric ICAM-1 molecules and four binding sites. In other preferred embodiments, the immunoglobulin heavy-chain portion of the chimeric molecule is derived from IgM and multivalent complexes containing ten or twelve molecules may be formed. In other preferred embodiments, in which the chimeric ICAM-1 molecule uses a chimeric immunoglobulin heavy-chain, the chimeric ICAM-1 molecule may form dimers or other higher order multivalent complexes through the domains from either IgA or IgM that are responsible for J chain binding. In other chimeric immunoglobulin molecules the portions of the immunoglobulin responsible for the disulfide bonding between the two immunoglobulin heavy-chains and/or the disulfide bonding between an immunoglobulin light-chain and heavy-chain may be placed in the chimeric immunoglobulin molecule to allow the formation of dimers or other high order multivalent complexes.

The present invention contemplates immunoadhesins containing a chimeric ICAM-1 molecule in which the immunoglobulin domains comprising the heavy chain are derived from different isotypes of either heavy or light chain immunoglobulins. One skilled in the art will understand that using molecular techniques, these domains can be substituted for a similar domain and thus produce an immunoglobulin that is a hybrid between two different immunoglobulin molecules. These chimeric immunoglobulins allow immunoadhesins containing secretory component to be constructed that contain a variety of different and desirable properties that are conferred by different immunoglobulin domains.

The present invention also contemplates chimeric ICAM-1 molecules in which the portion of the chimeric molecule derived from immunoglobulin, heavy or light J chain may contain less than an entire domain derived from a different immunoglobulin molecule. The same molecular techniques may be employed to produce such chimeric ICAM-1 molecules.

In preferred embodiments, the chimeric ICAM-1 molecules of the present invention contain at least the C_H1, C_H2, C_H3, domain of mouse or human IgA_1, IgA_2 or IgM. Other preferred embodiments of the present invention contain immunoglobulin domains that include at least the Cμ1, Cμ2, Cμ3, or Cμ4 domains of IgM.

Preferred chimeric ICAM-1 molecules contain domains from two different isotypes of human immunoglobulin. Preferred chimeric ICAM-1 molecules that include immunoglobulins that contain immunoglobulin domains including at least the C_H1, C_H2, or C_H3 of human IgG, IgG1, IgG2, IgG3, IgG4, IgA1, IgA2, IgE, or IgD. Other preferred immunoglobulins for use as part of chimeric ICAM-1 molecules include immunoglobulins that contain domains from at least the C_H1, C_H2, C_H3, or C_H4 domain of IgM or IgE. The present invention also contemplates chimeric ICAM-1 molecules that contain immunoglobulin domains derived from at least two different isotypes of mammalian immunoglobulins. Generally, any of the mammalian immunoglobulins can be used in the preferred embodiments, such as the following isotypes: any isotype of IgG, any isotype of IgA, IgE, IgD or IgM. The present invention also contemplates chimeric ICAM-1 molecules

derived from a species such as human, mouse or other mammals. In preferred embodiments, the chimeric ICAM-1 molecule contains the portion of IgA or IgM responsible for the association of J chain with the IgA and IgM. Thus, by using a chimeric immunoglobulin in the chimeric ICAM-1 molecule, the J chain may associate with a chimeric immunoglobulin that is predominantly of an isotype that does not bind J chain or secretory component.

The present invention also contemplates chimeric ICAM-1 molecules that contain immunoglobulin domains derived from two different isotypes of rodent or primate immunoglobulin. The isotypes of rodent or primate immunoglobulin are well known in the art. The chimeric ICAM-1 molecules of the present invention may contain immunoglobulin derived heavy chains that include at least one of the following immunoglobulin domains: the C_H1, C_H2, or C_H3 domains of a mouse IgG, IgG1, IgG2a, IgG2b, IgG3, IgA, IgE, or IgD; the C_H1, C_H2, C_H3 or C_H4 domain of mouse IgE or IgM; the C_H1 domain of a human IgG, IgG1, IgG2, IgG3, IgG4, IgA1, IgA2, or IgD; the C_H1, C_H2, C_H3, C_H4 domain of human IgM or IgE; the C_H1, C_H2, or C_H3 domain of an isotype of mammalian IgG, an isotype of IgA, IgE, or IgD; the C_H1, C_H2, C_H3 or C_H4 domain of a mammalian IgE or IgM; the C_H1, C_H2 , or C_H3 domain of an isotype of rodent IgG, IgA, IgE, or IgD; the C_H1, C_H2, C_H3 or C_H4 domain of a rodent IgE or IgM; the C_H1, C_H2, or C_H3 domain of an isotype of animal IgG, an isotype of IgA, IgE, or IgD; and the C_H1, C_H2, C_H3, or C_H4 domain of an animal IgE or IgM. The present invention also contemplates the replacement or addition of protein domains derived from molecules that are members of the immunoglobulin superfamily into the chimeric ICAM-1 molecules. The molecules that belong to the immunoglobulin superfamily have amino acid residue sequence and nucleic acid sequence homology to immunoglobulins. The molecules that are part of the immunoglobulin superfamily can be identified by amino acid or nucleic acid sequence homology. See, for example, p. 361 of Immunoglobulin Genes, Academic Press (1989).

In preferred embodiments of the present invention, the immunoadhesin is expressed by methods that generate an immunoadhesin having plant-specific glycosylation. It is well-known in the art that glycosylation is a major modification of proteins in plant cells

(Lerouge *et al.*, *Plant Molecular Biology* 38:31-48, 1998). Glycosylation of proteins also occurs in other cell types, including mammalian and insect cells. The glycosylation process starts in the endoplasmic reticulum by the co-translational transfer of a precursor oligosaccharide to specific residues of the nascent polypeptide chain. Processing of this oligosaccharide into different types of glycans, which differ in the types of residues present and the nature of the linkages between the residues, occurs in the secretory pathway as the glycoprotein moves from the endoplasmic reticulum to its final destination. One of skill in the art will understand that at the end of their maturation, proteins expressed in plants or plant cells have a different pattern of glycosylation than do proteins expressed in other types of cells, including mammalian cells and insect cells. Detailed studies characterizing plant-specific glycosylation and comparing it with glycosylation in other cell types have been performed, for example, in studies described by Cabanes-Macheteau *et al.*, *Glycobiology* 9(4):365-372 (1999), and Altmann, *Glycoconjugate J.* 14:643-646 (1997). These groups and others have shown that plant-specific glycosylation generates glycans that have xylose linked β(1,2) to mannose, but xylose is not linked β(1,2) to mannose as a result of glycosylation in mammalian and insect cells. Plant-specific glycosylation results in a fucose linked α(1,3) to the proximal GlcNAc, while glycosylation in mammalian cells results in a fucose linked α(1,6) to the proximal GlcNAc. Furthermore, plant-specific glycosylation does not result in the addition of a sialic acid to the terminus of the protein glycan, whereas in glycosylation in mammalian cells, sialic acid is added.

The immunoadhesin of the present invention that is glycosylated in a plant-specific manner can contain a chimeric ICAM-1 molecule that includes any combination of extracellular domains 1, 2, 3, 4, and 5 of the ICAM-1 molecule. FIG. 2B shows the amino acid sequence of the chimeric ICAM-1/IgA2 molecule (SEQ ID NO: 8) of the present invention, that contains all five domains of ICAM-1. The bolded N's represent asparagine residues to which oligosaccharide moieties are linked during glycosylation in plant cells, as well as mammalian and insect cells. One of skill in the art will know that the glycosylation sites are the tripeptide Asn-X-Ser/Thr where X can be any amino acid except proline and

aspartic acid (Kornfeld and Kornfeld, *Annu Rev Biochem* 54:631-664, 1985). It will therefore be known to one of skill in the art that which amino acids of the protein having plant-specific glycosylation would depend on which domains of ICAM-1 are present. Because the sequence and domain boundaries of ICAM-1 are known (see Staunton *et. al.*, *Cell* 52:925-933, 1988), it would be evident to one of skill in the art how to determine the plant-specific glycosylation sites on any potential combination of any of the five ICAM-1 domains.

In other preferred embodiments of the present invention, the immunoadhesin having plant-specific glycosylation and containing a chimeric ICAM-1 molecule having any combination of ICAM-1 extracellular domains 1, 2, 3, 4 and 5 further comprises a J chain and secretory component associated with said chimeric ICAM-1 molecule. As was true with respect to the chimeric ICAM-1 molecule, one of skill in the art will be able to identify the sites for plant-specific glycosylation in the J chain and secretory component sequences.

The present invention contemplates immunoadhesins having plant-specific glycosylation, that contain a chimeric ICAM-1 molecule in which the immunoglobulin heavy chain is selected from the group of IgA (SEQ ID NOS: 15-18), IgA_1 (SEQ ID NOS: 15-16), IgA_2 (SEQ ID NO: 17), IgG_1 (SEQ ID NOS: 19-20), IgG_2 (SEQ ID NOS: 21-22), IgG_3 (SEQ ID NOS: 23-24), IgG_4 (SEQ ID NOS: 25-26), IgM (SEQ ID NOS: 46-47), IgD (SEQ ID NO: 27-32 and 36-41-43), IgE (SEQ ID NOS: 44-45), and a chimeric immunoglobulin heavy chain. One of skill in the art will know that which of these immunoglobulin heavy chain sequences, or which combination of immunoglobulin heavy chain sequences are combined in a chimeric immunoglobulin heavy chain, will have an effect on the number and location of glycosylation sites in the chimeric ICAM-1 molecule of the immunoadhesin. As was true with respect to the chimeric ICAM-1 molecule, one of skill in the art will be able to identify the sites for plant-specific glycosylation in the immunoglobulin heavy chain sequences, including the various chimeric immunoglobulin heavy chain sequences that can be constructed.

Also provided herein are immunoadhesin functional derivatives. By "functional derivative" is meant a "chemical derivative," "fragment," or "variant," of the polypeptide or nucleic acid of the invention which retains at least a portion of the function of the protein, for example reactivity with an antibody specific for the protein, enzymatic activity or binding activity, which permits its utility in accordance with the present invention. It is well known in the art that due to the degeneracy of the genetic code numerous different nucleic acid sequences can code for the same amino acid sequence. It is also well known in the art that conservative changes in amino acid can be made to arrive at a protein or polypeptide that retains the functionality of the original. In both cases, all permutations are intended to be covered by this disclosure.

The derivatives may also be engineered according to routine methods to include an affinity purification tag such that large quantities and/or relatively pure or isolated quantities of immunoadhesin may be produced. Many different versions of tag exist that can be incorporated into one or more components of the immunoadhesin, preferably not destroying the desired binding activity of the immunoadhesin in the absence of tag. Such tags can be engineered as expressible encoded nucleic acid sequence fused with nucleic acid sequences encoding the immunoadhesins of the invention. The tags may further be engineered to be removable, *e.g.*, with a commercially available enzyme.

Further, it is possible to delete codons or to substitute one or more codons with codons other than degenerate codons to produce a structurally modified polypeptide, but one which has substantially the same utility activity as the polypeptide produced by the unmodified nucleic acid molecule. As recognized in the art, the two polypeptides can be functionally equivalent, as are the two nucleic acid molecules that give rise to their production, even though the differences between the nucleic acid molecules are not related to the degeneracy of the genetic code.

Manipulations of this sort, and post-production chemical derivatization may be implemented, *e.g.*, to improve stability, solubility, absorption, biological or therapeutic effect,

and/or biological half-life. Moieties capable of mediating such effects are disclosed, for example, in Remington's Pharmaceutical Sciences, 18th ed., Mack Publishing Co., Easton, PA (1990). A functional derivative intended to be within the scope of the present invention is a "variant" polypeptide which either lacks one or more amino acids or contains additional or substituted amino acids relative to the native polypeptide. The variant may be derived from a naturally occurring complex component by appropriately modifying the protein DNA coding sequence to add, remove, and/or to modify codons for one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence. It is understood that such variants having added, substituted and/or additional amino acids retain one or more characterizing portions of the native protein, as described above.

A functional derivative of a protein with deleted, inserted and/or substituted amino acid residues may be prepared using standard techniques well-known to those of ordinary skill in the art. For example, the modified components of the functional derivatives may be produced using site-directed mutagenesis techniques (as exemplified by Adelman *et. al.*, 1983, DNA 2:183) wherein nucleotides in the DNA coding sequence are modified such that a modified coding sequence is produced, and thereafter expressing this recombinant DNA in a prokaryotic or eukaryotic host cell, using techniques such as those described above. Alternatively, proteins with amino acid deletions, insertions and/or substitutions may be conveniently prepared by direct chemical synthesis, using methods well-known in the art. The functional derivatives of the proteins typically exhibit the same qualitative biological activity as the native proteins.

In addition, the immunoadhesins of the invention may be not just modified ICAM-1/Ig immunoadhesins, but may also embrace other native ICAM family members, isotypes, and/or other homologous amino acid sequences, *e.g.*, human, primate, rodent, canine, feline, bovine, avian, etc. Furthermore, the Ig type used in the immunoadhesins can vary, *e.g.*, may assume a different Ig family member identity, within or without a given species. ICAMs and Igs are diverse and have well-known sequences that one of ordinary skill can exploit to create different immunoadhesins having more or less different utility in a given organism to undergo

treatment. An illustrative, nonexhaustive list of examples of molecules having ICAM-1 homology that can be used to create other immunoadhesins include those in the following table.

Table 3

ACCESSION NO.	ICAM NAME	SPECIES
NP 000192	Intercellular Adhesion Molecule-1 (CD54)	Homo sapiens
AAH03097	Intercellular Adhesion Molecule ICAM-2	Homo sapiens
NP 002153	Intercellular Adhesion Molecule 3 Precursor	Homo sapiens
BAB20325	TCAM-1	Homo sapiens
NP 003250	Intercellular Adhesion Molecule 5 (Telencephalin)	Homo sapiens
NM 007164	Mucosal Vascular Address in Cell Adhesion Molecule (MADCAM1)	Homo sapiens
NM 001078	Vascular Cell Adhesion Molecule 1 (VCAM1)	Homo sapiens
AAA37875	MALA-2	Mus musculus
AAA37876	Intercellular Adhesion Molecule-1 Precursor	Mus musculus
AAG30280	Intracellular Adhesion Molecule 1	Cricetulus griseus
AAB39264	Intercellular Adhesion Molecule-3	Bos taurus
AAF80287	Intercellular Adhesion Molecule-1 Precursor	Sus scrofa
AAA18478	Telecephalin	Oryctolagus cuniculus

ACCESSION NO.	ICAM NAME	SPECIES
NP 032345	Intercellular Adhesion Molecule 5, telencephalin	Mus musculus
BAB41106	Cell adhesion molecule TCAM-1	Mus musculus
NP 067705	Testicular Cell Adhesion Molecule 1	Rattus norvegicus
AAG35584	Nectin-Like Protein 1	Mus musculus
AAC18956	CD22 Protein	Homo sapiens
AAA35415	Intercellular Adhesion Molecule 1	Pan troglodytes
AAA83206	89 kDa Protein	Mus musculus
AAA92551	Intercellular Adhesion Molecule-1	Canis familiaris
AAB06749	Intercellular Adhesion Molecule-1	Bos taurus
AAD13617	Intercellular Adhesion Molecule-1 Precursor	Ovis aries
NP 037099	Intercellular Adhesion Molecule-1	Rattus norvegicus
AAE22202	ICAM-4	Rattus norvegicus
AAA60392	cell surface glycoprotein	Homo sapiens
AAF91086	nephrin	Rattus norvegicus
AAF91087	nephrin	Mus musculus

Likewise, numerous heavy chain constant regions of different Ig molecules, both in humans and other species, are known that can be substituted in for those specific Ig regions of the chimeras described herein.

C. Vectors, Cells and Plants Containing Immunoadhesins

The present invention also contemplates expression and cloning vectors, cells and plants containing the immunoadhesins of the present invention. Technology for isolating the genes encoding the various portions of the immunoadhesions are well-known to one of skill in the art and can be applied to insert the various required genes into expression vectors and cloning vectors such as those vectors can be introduced into cells and into transgenic plants.

The present invention contemplates a method of assembling an immunoadhesin comprising the steps of: introducing into an organism a DNA segment encoding a chimeric ICAM-1 molecule, immunoglobulin J chain, and introducing into the same organism a DNA encoding a secretory component. The preferred secretory component contains at least a segment of the amino acid residues 1 to residue about 606 of the human polyimmunoglobulin receptor (pIgR) amino acid residue sequence or analogous amino acid residues from other species (Mostov, *Ann Dev. Immu.* 12:63-84 1994).

The present invention contemplates eukaryotic cells, including plant cells, containing immunoadhesins of the present invention. The present invention also contemplates plant cells that contain nucleotide sequences encoding the various components of the immunoadhesin of the present invention. One skilled in the art will understand that the nucleotide sequences that encode the secretory component protection protein and the chimeric ICAM-1 molecule and J chain will typically be operably linked to a promoter and present as part of an expression vector or cassette. Typically, if the eukaryotic cell used is a plant cell than the promoter used will be a promoter that is able to operate in a plant cell. After the chimeric ICAM-1 genes, secretory component genes and J chain genes are isolated, they are typically operatively linked to a transcriptional promoter in an expression vector. The present invention also contemplates expression vectors containing a nucleotide sequence encoding a chimeric

ICAM-1 molecule which has been operatively linked to a regulatory sequence for expression. These expression vectors place the nucleotide sequence to be expressed in a particular cell 3' of a promoter sequence which causes the nucleotide sequence to be transcribed and expressed. The expression vector may also contain various enhancer sequences which improve the efficiency of this transcription. In addition, such sequences as terminators, polyadenylation (poly A) sites and other 3' end processing signals may be included to enhance the amount of nucleotide sequence transcribed within a particular cell.

Expression of the components in the organism of choice can be derived from an independently replicating plasmid, or from a permanent component of the chromosome, or from any piece of DNA which may transiently give rise to transcripts encoding the components. Organisms suitable for transformation can be either prokaryotic or eukaryotic. Introduction of the components of the complex can be by direct DNA transformation, by biolistic delivery into the organism, or mediated by another organism as for example by the action of recombinant Agrobacterium on plant cells. Expression of proteins in transgenic organisms usually requires co-introduction of an appropriate promoter element and polyadenylation signal. In one embodiment of the invention, the promoter element potentially results in the constitutive expression of the components in all of the cells of a plant. Constitutive expression occurring in most or all of the cells will ensure that precursors can occupy the same cellular endomembrane system as might be required for assembly to occur.

Expression vectors compatible with the host cells, preferably those compatible with plant cells are used to express the genes of the present invention. Typical expression vectors useful for expression of genes in plants are well known in the art and include vectors derived from the tumor-inducing (Ti) plasmid of Agrobacterium tumefaciens described by Rogers *et al.*, *Meth. in Enzymol.*, 153:253-277 (1987). However, several other expression vector systems are known to function in plants. See for example, Verma *et al.*, PCT Publication No. WO87/00551; and Cocking and Davey, *Science*, 236:1259-1262 (1987).

The expression vectors described above contain expression control elements including the promoter. The genes to be expressed are operatively linked to the expression vector to allow the promoter sequence to direct RNA polymerase binding and synthesis of the desired polypeptide coding gene. Useful in expressing the genes are promoters which are inducible, viral, synthetic, constitutive, and regulated. The choice of which expression vector is used and ultimately to which promoter a nucleotide sequence encoding part of the immunoadhesin of the present invention is operatively linked depends directly, as is well known in the art, on the functional properties desired, *e.g.* the location and timing of protein expression, and the host cell to be transformed, these being limitations inherent in the art of constructing recombinant DNA molecules. However, an expression vector useful in practicing the present invention is at least capable of directing the replication, and preferably also the expression of the polypeptide coding gene included in the DNA segment to which it is operatively linked.

In preferred embodiments, the expression vector used to express the genes includes a selection marker that is effective in a plant cell, preferably a drug resistance selection marker. A preferred drug resistance marker is the gene whose expression results in kanamycin resistance, i.e., the chimeric gene containing the nopaline synthase promoter, Tn5 neomycin phosphotransferase II and nopaline synthase 3' nontranslated region described by Rogers *et al.*, in *Methods For Plant Molecular Biology*, a Weissbach and H. Weissbach, eds., Academic Press Inc., San Diego, Calif. (1988). A useful plant expression vector is commercially available from Pharmacia, Piscataway, N.J. Expression vectors and promoters for expressing foreign proteins in plants have been described in U.S. Pat. Nos. 5,188,642; 5,349,124; 5,352,605, and 5,034,322 which are hereby incorporated by reference.

A variety of methods have been developed to operatively link DNA to vectors via complementary cohesive termini. For instance, complementary homopolymer tracks can be added to the DNA segment to be inserted into the vector DNA. The vector and DNA segment are then joined by hydrogen bonding between the complementary homopolymeric tails to form recombinant DNA molecules. Alternatively, synthetic linkers containing one or more restriction endonuclease sites can be used to join the DNA segment to the expression vector.

The synthetic linkers are attached to blunt-ended DNA segments by incubating the blunt-ended DNA segments with a large excess of synthetic linker molecules in the presence of an enzyme that is able to catalyze the ligation of blunt-ended DNA molecules, such as bacteriophase T4 DNA ligase. Thus, the products of the reaction are DNA segments carrying synthetic linker sequences at their ends. These DNA segments are then cleaved with the appropriate restriction endonuclease and ligated into an expression vector that has been cleaved with an enzyme that produces termini compatible with those of the synthetic linker. Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of sources including New England BioLabs, Beverly, Mass.

The nucleotide sequences encoding the secretory component, J chain, the chimeric ICAM-1 molecules of the present invention are introduced into the same plant cell either directly or by introducing each of the components into a plant cell and regenerating a plant and cross-hybridizing the various components to produce the final plant cell containing all the required components.

Any method may be used to introduce the nucleotide sequences encoding the components of the immunoadhesins of the present invention into a eukaryotic cell. For example, methods for introducing genes into plants include *Agrobacterium*-mediated plant transformation, protoplast transformation, gene transfer into pollen, injection into reproductive organs and injection into immature embryos. Each of these methods has distinct advantages and disadvantages. Thus, one particular method of introducing genes into a particular eukaryotic cell or plant species may not necessarily be the most effective for another eukaryotic cell or plant species.

Agrobacterium tumefaciens-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, bypassing the need for regeneration of an intact plant from a protoplast. The use of *Agrobacterium*-mediated expression vectors to introduce DNA into plant cells is well known in the art. See, for example, the methods described by Fraley *et al.*, *Biotechnology*, 3:629

(1985) and Rogers *et al.*, *Methods in Enzymology*, 153:253-277 (1987). Further, the integration of the Ti-DNA is a relatively precise process resulting in few rearrangements. The region of DNA to be transferred is defined by the border sequences and intervening DNA is usually inserted into the plant genome as described by Spielmann *et al.*, *Mol. Gen. Genet.*, 205:34 (1986) and Jorgensen *et al.*, *Mol. Gen. Genet.*, 207:471 (1987). Modern *Agrobacterium* transformation vectors are capable of replication in *Escherichia* coli as well as *Agrobacterium*, allowing for convenient manipulations as described by Klee *et al.*, in *Plant DNA Infectious Agents*, T. Hohn and J. Schell, eds., *Springer-Verlag*, New York, pp. 179-203 (1985). Further recent technological advances in vectors for *Agrobacterium*-mediated gene transfer have improved the arrangement of genes and restriction sites in the vectors to facilitate construction of vectors capable of expressing various polypeptide coding genes. The vectors described by Rogers *et al.*, *Methods in Enzymology*, 153:253 (1987), have convenient multi-linker regions flanked by a promoter and a polyadenylation site for direct expression of inserted polypeptide coding genes and are suitable for present purposes.

In those plant species where *Agrobacterium*-mediated transformation is efficient, it is the method of choice because of the facile and defined nature of the gene transfer. *Agrobacterium*-mediated transformation of leaf disks and other tissues appears to be limited to plant species that *Agrobacterium* tumefaciens naturally infects. Thus, *Agrobacterium*-mediated transformation is most efficient in dicotyledonous plants.

Few monocots appear to be natural hosts for *Agrobacterium*, although transgenic plants have been produced in asparagus using *Agrobacterium* vectors as described by Bytebier *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 84:5345 (1987). Therefore, commercially important cereal grains such as rice, corn, and wheat must be transformed using alternative methods. Transformation of plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments. See, for example, Potrykus *et al.*, *Mol. Gen. Genet.*, 199:183 (1985); Lorz *et al.*, *Mol. Gen. Genet.*, 199:178 (1985); Fromm *et al.*, *Nature*, 319:791

(1986); Uchimiya *et al.*, *Mol. Gen. Genet.*, 204:204 (1986); Callis *et al.*, *Genes and Development*, 1:1183 (1987); and Marcotte *et al.*, *Nature*, 335:454 (1988).

Application of these methods to different plant species depends upon the ability to regenerate that particular plant species from protoplasts. Illustrative methods for the regeneration of cereals from protoplasts are described in Fujimura *et al.*, *Plant Tissue Culture Letters*, 2:74 (1985); Toriyama *et al.*, *Theor Appl. Genet.*, 73:16 (1986); Yamada *et al.*, *Plant Cell Rep.*, 4:85 (1986); Abdullah *et al.*, *Biotechnology*, 4:1087 (1986).

To transform plant species that cannot be successfully regenerated from protoplasts, other ways to introduce DNA into intact cells or tissues can be utilized. For example, regeneration of cereals from immature embryos or explants can be effected as described by Vasil, *Biotechnology*, 6:397 (1988). In addition, "particle gun" or high-velocity microprojectile technology can be utilized. Using such technology, DNA is carried through the cell wall and into the cytoplasm on the surface of small (0.525 μm) metal particles that have been accelerated to speeds of one to several hundred meters per second as described in Klein *et al.*, *Nature*, 327:70 (1987); Klein *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 85:8502 (1988); and McCabe *et al.*, *Biotechnology*, 6:923 (1988). The metal particles penetrate through several layers of cells and thus allow the transformation of cells within tissue explants. Metal particles have been used to successfully transform corn cells and to produce fertile, stably transformed tobacco and soybean plants. Transformation of tissue explants eliminates the need for passage through a protoplast stage and thus speeds the production of transgenic plants.

DNA can also be introduced into plants by direct DNA transfer into pollen as described by Zhou *et al.*, *Methods in Enzymology*, 101:433 (1983); D. Hess, *Intern Rev. Cytol.*, 107:367 (1987); Luo *et al.*, *Plant Mol. Biol. Reporter*, 6:165 (1988). Expression of polypeptide coding genes can be obtained by injection of the DNA into reproductive organs of a plant as described by Pena *et al.*, *Nature*, 325:274 (1987). DNA can also be injected directly into the cells of immature embryos and the rehydration of desiccated embryos as

described by Neuhaus *et al.*, *Theor. Appl. Genet.*, 75:30 (1987); and Benbrook *et al.*, in *Proceedings Bio Expo* 1986, Butterworth, Stoneham, Mass., pp. 27-54 (1986).

The regeneration of plants from either single plant protoplasts or various explants is well known in the art. See, for example, Methods for Plant Molecular Biology, A. Weissbach and H. Weissbach, eds., Academic Press, Inc., San Diego, Calif. (1988). This regeneration and growth process includes the steps of selection of transformant cells and shoots, rooting the transformant shoots and growth of the plantlets in soil.

The regeneration of plants containing the foreign gene introduced by *Agrobacterium* tumefaciens from leaf explants can be achieved as described by Horsch *et al.*, *Science*, 227:1229-1231 (1985). In this procedure, transformants are grown in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant species being transformed as described by Fraley *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 80:4803 (1983). This procedure typically produces shoots within two to four weeks and these transformant shoots are then transferred to an appropriate root-inducing medium containing the selective agent and an antibiotic to prevent bacterial growth. Transformant shoots that rooted in the presence of the selective agent to form plantlets are then transplanted to soil to allow the production of roots. These procedures will vary depending upon the particular plant species employed, such variations being well known in the art.

The immunoadhesins of the present invention may be produced in any plant cell including plant cells derived from plants that are dicotyledonous or monocotyledonous, solanaceous, alfalfa, legumes, or tobacco.

Transgenic plants of the present invention can be produced from any sexually crossable plant species that can be transformed using any method known to those skilled in the art. Useful plant species are dicotyledons including tobacco, tomato, the legumes, alfalfa, oaks, and maples; monocotyledons including grasses, corn, grains, oats, wheat, and barley; and lower plants including gymnosperms, conifers, horsetails, club mosses, liverworts, hornworts, mosses, algaes, gametophytes, sporophytes or pteridophytes.

The present invention also contemplates expressing the immunoadhesins within eukaryotic cells including mammalian cells. One of skill in the art will understand the various systems available for expression of the immunoadhesin in mammalian cells and can readily modify those system to express the immunoadhesions and chimeric ICAM-1 molecules of the present invention in those cells. In preferred embodiments, the chimeric ICAM-1, J chain and secretory component molecules of the present invention are placed in a vector pCDM8 which has been previously described by Aruffo, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 84:8573-8577 (1987). The use of the PCDM8 construct is by no means unique and numerous other systems are available that do not utilize the cog cell expression system. For example, the following United States Patents describe useful eukaryotic expression systems that may be used with the chimeric ICAM-1 and other molecules of the immunoadhesin.

D. Compositions Containing Immunoadhesins

The present invention also contemplates compositions containing an immunoadhesin of the present invention together with plant macromolecules or material. Typically these plant macromolecules or plant materials are derived from any plant useful in the present invention. The plant macromolecules are present together with an immunoadhesin of the present invention for example, in a plant cell, in an extract of a plant cell, or in a plant. Typical plant macromolecules associated with the immunoadhesin of the present invention in a composition are ribulose bisphosphate carboxylase, light harvesting complex pigments (LHCP), secondary metabolites or chlorophyll. The compositions of the present invention have plant material or plant macromolecules in a concentration of between 0.01% and 99% mass excluding water. Other compositions include compositions having the immunoadhesins of the present invention present at a concentration of between 1% and 99% mass excluding water. Other compositions include immunoadhesins at a concentration of 50% to 90% mass excluding water.

The compositions of the present invention may contain plant macromolecules at a concentration of between 0.1% and 5% mass excluding water. Typically the mass present in the composition will consist of plant macromolecules and immunoadhesins of the present

invention. When the immunoadhesins of the present invention are present at a higher or lower concentration the concentration of plant macromolecules present in the composition will vary inversely. In other embodiments the composition of plant macromolecules are present in a concentration of between 0.12% and 1% mass excluding water.

The present invention contemplates a composition of matter comprising all or part of the following: a chimeric ICAM-1 molecule, a J chain or a secretory component. These components form a complex and are associated as was previously described. Typically, the composition also contains molecules derived from a plant. This composition may also be obtained after an extraction process yielding functional immunoadhesin and plant-derived molecules.

The extraction method comprises the steps of applying a force to a plant containing the complex whereby the apoplastic compartment of the plant is ruptured releasing said complex. The force involves shearing as the primary method of releasing the apoplastic liquid.

The whole plant or plant extract contains an admixture of immunoadhesin and various other macromolecules of the plant. Among the macromolecules contained in the admixture is ribulose bisphosphate carboxylase (RuBisCo) or fragments of RuBisCo. Another macromolecule is LHCP. Another molecule is chlorophyll.

Other useful methods for preparing compositions containing immunoadhesins having chimeric ICAM-1 molecule include extraction with various solvents and application of vacuum to the plant material. The compositions of the present invention may contain plant macromolecules in a concentration of between about 0.1% and 5% mass excluding water.

The present invention also contemplates therapeutic compositions which may be used in the treatment of a patient or animal. Administration of the therapeutic composition can be before or after extraction from the plant or other transgenic organism. Once extracted the immunoadhesins may also be further purified by conventional techniques such as size

exclusion, ion exchange, or affinity chromatography. Plant molecules may be co-administered with the complex.

The present invention also contemplates that the relative proportion of plant-derived molecules and animal-derived molecules can vary. Quantities of specific plant proteins, such as RuBisCo or chlorophyll may be as little as 0.01% of the mass or as much as 99.9% of the mass of the extract, excluding water.

The present invention also contemplates the direct use of the therapeutic plant extract containing immunoadhesins without any further purification of the specific therapeutic component. Administration may be by topical application, oral ingestion, nasal spray or any other method appropriate for delivering the antibody to the mucosal target pathogen.

E. Pharmaceutical Compositions, Formulations, And Routes Of Administration

The immunoadhesins described herein can be administered to a patient, preferably in the form of a suitable pharmaceutical composition. Such composition may include components in addition to, or in lieu of, those described above. The composition preferably exhibits either or both of a therapeutic and prophylactic property when administered. The preparation of such compositions can be done according to routine methodologies in the art, and may assume any of a variety of forms, e.g., liquid solutions, suspensions or emulsifications, and solid forms suitable for inclusion in a liquid prior to ingestion. Techniques for the formulation and administration of polypeptides and proteins may be found in *Remington's Pharmaceutical Sciences*, Mack Publishing Co., Easton, PA, latest edition. Using these procedures, one of ordinary skill can utilize the immunoadhesins of the invention to achieve success without undue experimentation.

1. Administration Routes

Suitable routes of administration for the invention include, *e.g.*, oral, nasal, inhalation, intraocular, phanyngeal, bronchial, transmucosal, or intestinal administration. Alternatively,

one may administer the compound in a local manner, *e.g.*, via injection or other application of the compound to a preferred site of action.

2. Formulations

The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. One or more physiologically acceptable carriers comprising excipients and/or other auxiliaries can be used to facilitate processing of the active compounds into pharmaceutical preparations. Proper formulation is dependent upon the particular route of administration chosen.

For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Suitable carriers include excipients such as, e.g., fillers such as sugars, including lactose, sucrose, mannitol, and/or sorbitol; cellulose preparations such as, e.g., maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl- cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration.

For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention may be conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of *e.g.* gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

In addition, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols.

Pharmaceutically compatible salts may be formed with many acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, citric, etc. Salts tend to be more soluble in aqueous or other protonic solvents that are the corresponding free base forms. In solutions, manipulation of pH is also routinely employed for optimizing desired properties.

3. Determining Effective Dosages and Dosage Regimens

Pharmaceutical compositions suitable for use in the present invention include compositions where the active ingredients are contained in an amount effective to achieve an intended purpose, *e.g.*, a therapeutic and/or prophylactic use. A pharmaceutically effective

amount means an amount of compound effective to prevent, alleviate or ameliorate symptoms of disease or prolong the survival of the subject being treated. Determination of a pharmaceutically effective amount is well within the capability of those skilled in the art, and will typically assume an amount of between about 0.5 μg/kg/day and about 500g/kg/day, with individual dosages typically comprising between about 1 nanogram and several grams of immunoadhesin.

For any compound used in the methods of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. For example, varying dosages can be administered to different animals or cell cultures and compared for effect. In this way, one can identify a desired concentration range, and prepare and administer such amount accordingly. Optimization is routine for one of ordinary skill in the art.

The person of skill, in addition to considering pharmaceutical efficacy, also considers toxicity according to standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD50 and ED50. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (See *e.g.*, Fingl *et al.*, 1975, in "*The Pharmacological Basis of Therapeutics*," Ch. 1 p.1).

Dosage amount and frequency may be adjusted to provide mucosal levels of immunadhesin sufficient to maintain or provide a pharmaceutical effect, e.g., therapeutic and/or prophylactic. The minimal effective concentration (MEC) will vary for each

immunadhesin and immunoadhesin formulation, but can be estimated from *in vitro* and/or *in vivo* data. Dosages necessary to achieve MEC will depend on individual characteristics and route of administration. However, assays as described herein can be used to determine mucosal concentrations, which can then be further optimized in amount and precise formulation.

Dosage intervals can also be determined using MEC value. Compounds can be administered using a regimen which maintains mucosal levels above the MEC for 10-90% of the time, 30-90% of the time, or, most preferably, 50-90% of the time.

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. The pack or dispenser may also be accompanied with a notice associated with the container in form prescribed by a governmental agency regulating the manufacture, use, or sale of pharmaceuticals, which notice is reflective of approval by the agency of the form of the immunoadhesin for human or veterinary administration. Such notice, for example, may be the labeling approved by the U.S. Food and Drug Administration for prescription drugs, or the approved product insert. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition, *e.g.* treatment or prophylaxis of a disease mediated by host organism/patient ICAM molecules.

F. Methods of Treatment and Prevention of ICAM-mediated Afflictions

A patient in need of therapeutic and/or prophylactic immunoadhesin chimeras of the invention, *e.g.*, to counter rhinovirus infection and/or symptoms such as occur with colds, can be administered a pharmaceutically effective amount of desired immunoadhesin, preferably as part of a pharmaceutical composition determined, produced, and administered as described above. These formulations and delivery modalities can vary widely. Described following are

preliminary procedures that can be used to deduce effective amounts and toxicity, and which can then be conveniently used to determine treatment and prophylaxis parameters and regimens, both in humans and other animals. These procedures are illustrative only and are not intended to be limiting of the invention. Further, these procedures are routine for one of ordinary skill in the art.

1. Ability of the Immunoadhesin to Reduce Rhinovirus Infectivity in Humans: Dose Escalation Tolerance Study

Immunoadhesins of the invention may be tested, *e.g.*, using randomized controlled trials to determine the effect of administration, e.g., intranasal, of immunoadhesin on infection. Other administration routes can be used. Various assays exist that can be used to monitor effect, e.g., IL-8 response assays assays that evaluate illness symptoms, e.g., cold symptoms caused by rhinovirus infection. These studies can evaluate the extent to which an immunoadhesin taken by a patient subjects can prevent or treat rhinovirus infection. For example, healthy or unhealthy subjects can be administered the immunoadhesin and evalutated over a time course, e.g., in tandem with rhinovirus inoculation and/or illness progression. The clinical protocols used may be based on protocols previously used in evaluation of a recombinant soluble ICAM-1 molecule for efficacy against rhinovirus infection, or modifications thereto (Turner, *et. al.*, *JAMA* 281:1797-804, 1999).

Male and female subjects of any species, age, health, or disease state can be evaluated. The subjects may exhibit a serum neutralizing antibody titer in advance of study, which titer may fluctuate in response to infection and immunoadhesin administration.

The immunoadhesin of the present invention may be formulated as a buffered saline with varying amounts of immunoadhesin within and administered at various intervals to a patient. Single ascending dose and multiple ascending dose studies can be used to evaluate the safety of the immunoadhesin. In each case, one or more subjects may be evaluated at each dosage level, some receiving the immunoadhesin, and one or more optionally receiving

placebo. In either study, multiple dosage levels may be evaluated. Dosage levels can vary, but are typically in the nanogram to gram range.

Dosages may be administered over seconds, minutes, hours, weeks, and months, and evaluated for toxicity and/or pharmaceutical effect.

Safety and toxicity may be assessed, e.g., by visual examination of the nasal mucosa for signs of irritation or inflammation. Blood safety evaluations can also be employed according to routine methods and using sensitive assays such as ELISA to determine various blood components, including circulating immunoadhesin and rhinovirus quantities. Naval lavage testing may similarly be done according to routine methodologies.

Routine statistical analyses and calculations may be employed to determine efficacy and toxicity predicted over time courses for single patients and/or for populations of patient-recipients..

Challenge studies as well known in the art can be used to demonstrate that treatment protects against clinical colds and/or reduces cold symptoms after viral challenge, and using commercially available starting materials such virus, cells, and animals. *See, e.g.*, Gwaltney, *et. al., Prog. Med. Virol.* 39:256-263, 1992.

The following examples illustrate the disclosed invention. These examples in no way limit the scope of the claimed invention.

EXAMPLES

1. Construction of Immunoadhesin Expression Cassettes

A cassette encoding ICAM-1 extracellular domains D1 through D5 was prepared by PCR cloning. Specifically, a fragment containing all five extracellular Ig-like domains of ICAM-1 was amplified from plasmid pCDIC1-5D/IgA (insert Martin, *et al.* reference) using the following oligonucleotide primers:

5'-TCTGTTCCCAGGAACTAGTTTGGCACAGACATCTGTGTCCCCCTCAAAAGTC-3' (SEQ ID NO: 6)

5'-CATACCGGGGACTAGTCACATTCACGGTCACCTCGCGG-3' (SEQ ID NO: 7)

These two primers were designed to introduce SpeI sites at the 5' and 3' ends of the PCR fragment (underlined nucleotides). PCR was performed with Pfu polymerase (Stratagene) to reduce accumulation of errors. The PCR fragment was cloned into the vector PCRScript (Stratagene), and sequenced before fusing to the human IgA2 cassettes (with and without SEKDEL at the carboxy-terminus).

Constructs for the expression in plants of human J chain and secretory component, as well as a human IgA2 heavy chain, were developed. A heavy chain expression cassette vector was made and called pSSpHuA2 (See **FIG. 1**). It contains sequence encoding a bean legumin signal peptide (Baumlein *et al.*, *Nucleic Acids Res.* 14 (6), 2707-2720, 1986). The sequence of bean legumin is provided as GenBank Accession No. X03677, and the sequence of the bean legumin signal peptide is SEQ ID NO: 10 (also see Fig. 8) and the IgA2m(2) constant region with SpeI and SacI sites in between, and the SuperMas promoter for driving the expression of a signal peptide and the constant regions of the human IgA2m(2) heavy-chain.

The amplified DNAs encoding the first five domains of human ICAM-1, and the Fc region of human IgA2m(2) were fused in a plant-expression cassette to make a chimeric

ICAM-1 molecule expression construct, shown in FIG. 2A. This was done by cloning the fragment encoding the five extracellular domains of ICAM-1 into vector pSSPHuA2 to generate pSSPICAMHuA2. The convenient restriction sites (5' SpeI and 3' Spe I) allowed the amplified fragment to be inserted between the signal peptide and the Cα1 domain. In the resulting construct, expression of the chimeric ICAM-1 molecule is under the control of the constitutive promoter "superMAS" (Ni *et. al.*, 1995) and the nos 3' terminator region.

The resulting chimeric ICAM-1 molecule construct contains no variable region. Upon translation of the mRNA, signal peptide cleavage is predicted to deposit the ICAM-1-heavy chain fusion into the plant cell's endoplasmic reticulum (ER). DNA encoding an ER retention signal (RSEKDEL, SEQ ID NO: 5) was appended to the 3' end of the heavy-chain in order to boost the expression level of the construct. The amino acid sequence SEKDEL (SEQ ID NO: 4) is the consensus signal sequence for retention of proteins in the endoplasmic reticulum in plant cells. This sequence has been shown to enhance accumulation levels of antibodies in plants (Schouten *et al., Plant Molecular Biology* 30:781-793,1996). The amino acid sequence of the chimeric ICAM-1 molecule construct is shown in FIG. 2B. The DNA sequence and translational frame of the construct was verified before it was used for particle bombardment.

It has been shown recently that assembly of J chain with IgA takes place in the Golgi apparatus (Yoo *et al., J. Biol. Chem.* 274:33771-33777, 1999), and so constructions of heavy chain without SEKDEL have been made as well. The ICAM-1 fragment was cloned into an expression cassette containing the IgA2m(2) constant region without SEKDEL.

2. Expression of Assembled Immunoadhesin in Plants

A. Immunoadhesin Expression Vectors

The plasmid pSSPICAMHuA2 (SEQ ID NO:9 and FIG. 8) is 6313 bp in length. Nucleotides 49-1165 represent the Superpromoter (Ni *et al., Plant Journal* 7:661-676, 1995). Nucleotides 1166-3662 comprise a sequence encoding a human ICAM-1/human IgA2m(2) constant hybrid with linker sequences. A consensus Kozak sequence (Kozak, *Cell* 44(2):283-92, 1986) is included (nt 1186-1192) to enhance translation initiation, as well as the signal

peptide from V. faba legumin (nt 1189-1257; Bäumlein *et al., Nucleic Acids Reg.* 14(6):2707-2720 (1986). The sequence of the human IgA2m(2) constant region (nt 3663-3633) has been previously published (Chintalacharuvu, *et al., J. Imm.* 152: 5299-5304, 1994). A sequence encoding the endoplasmic reticulum retention signal SEKDEL is appended to the end of the heavy Chain (nt 3634-3654). Nucleotides 3663-3933 derive from the nopaline synthase 3' end (transcription termination and polyadenlyation signal; Depicker et al., 1982). The remainder of the plasmid derives from the vector pSP72 (Promega).

The plasmid pSHuJ (SEQ ID NO: 11 and FIG. 8) is 4283 bp in length. Nucleotides 14-1136 represent the Superpromoter (Ni *et al., Plant Journal* 7:661-676, 1995) and nucleotides 1137-1648 are shown in FIG. 8 and comprise a sequence encoding the human J Chain including the native signal peptide (Max and Korsmeyer, *J Imm.* 152:5299-5304, 1985) along with linker sequences. A consensus Kozak sequence (Kozak, *Cell* 44(2):283-92, 1986) is included (nt 1162-1168) to enhance translation initiation. Nucleotides 1649-1902 derive from the nopaline synthase 3' end (transcription termination and polyadenlyation signal; Depicker *et al., J Mol Appl Genet* 1(6):561-73, 1982). The remainder of the plasmid derives from the vector pSP72 (Promega).

The plasmid pSHuSC (SEQ ID NO:12 and FIG. 8) is 5650 bp in length. Nucleotides 13-1136 are derived from the Superpromoter (Ni *et al., Plant Journal* 7:661-676, 1995), and nucleotides 1137-2981 comprise a sequence encoding the human Secretory Component including the native signal peptide (Krajci, *et al., Biochem. and Biophys. Res. Comm* 158:783, 1994) along with linker sequences. A consensus Kozak sequence (Kozak, *Cell* 44(2):283-92, 1986) is included (nt 1151-1157) to enhance translation initiation. Nucleotides 2982-3236 derive from the nopaline synthase 3' end, providing a transcription termination and polyadenlyation signal, described in Depicker *et al., J Mol Appl Genet* 1(6):561-73 (1982). The remainder of the plasmid derives from the vector pSP72 (Promega).

The plasmid pBMSP-1 (SEQ ID NO:13 and FIG. 8) is derived from pGPTV-KAN. Becker *et al.,* in Plant Molecular Biology 20, 1195-1197, (1992), describe new plant binary

vectors with selectable markers located proximal to the left T-DNA border, and the sequences outside of the left and right borders. Nucleotides 18-187 of pBMSP-1 represent the right T-DNA border, and nucleotides 1811-775 represent the superMAS promoter. Nucleotides 2393-2663 represent the NOS promoter (Depicker *et al.*, J *Mol Appl Genet* 1(6):561-73, 1982), nucleotides 2698-3492 encode the NPTII gene (conferring resistance to kanamycin), and nucleotides 3511-3733 are the polyadenylation signal from A. tumefaciens gene 7 (Gielen *et al., Embo* J 3:835-46, 1984). Nucleotides 1768-976 encode the NPTII gene, and nucleotides 4317-4464 represent the left T-DNA border.

The plasmid pBMSP-1spJSC (SEQ ID NO:14 and FIG. 8) is a derivative of pBMSP, containing both J and SC under control of superpromoter. In this plasmid, nucleotides 1-149 represent the left T-DNA border. Nucleotides 955-733 are the polyadenylation signal from A. tumefaciens gene, nucleotides 1768-976 encode the NPTII gene (conferring resistance to kanamycin), and nucleotides 2073-1803 represent the NOS promoter. Nucleotides 2635-3768 represent the superMAS promoter, nucleotides 3774-5595 encode the Human Secretory component, and nucleotides 5603-5857 represent the NOS polyadenylation signal. Nucleotides 5880-6991 represent the superMAS promoter, nucleotides 7007-7490 encode the Human Joining Chain, and nucleotides 7504-7757 represent the NOS polyadenylation signal. Nucleotides 7886-8057 represent the right T-DNA border.

The plasmid pGPTV-HPT, encoding the enzyme conferring hygromycin resistance, is available commercially from the Max-Planck-Institut für Züchtungsforschung (Germany). It is described by Becker in Plant Molecular Biology 20, 1195-1197 (1992).

B. Plant Transformation and Immunoadhesin Expression in Plants

The expression cassettes described above were used to produce the assembled immunoadhesin in plants. Plasmids pSSPICAMHuA2, pSHuJ, pSHuSC and pBMSP-1were co-bombarded into tobacco leaf tissue (*N. tabacum* cultivar Xanthi) and transformed microcalli were selected on nutrient agar in the presence of kanamycin. Individual microcalli,

indicative of independent transformation events, were dissected from the parent tissue and propagated on nutrient agar with kanamycin.

The callus tissues were screened for transgene expression. Callus #7132 was shown to express a chimeric ICAM-1 immunoadhesin and J chain by immunoblotting and PCR (data not shown). This callus did not possess DNA encoding the SC. The callus grew well in culture and, upon accumulation of sufficient mass, #7132 was bombarded again, this time with two of the plasmids described above, PBMSP-1 SpJSC, containing expression cassettes for both the J chain and SC and pGPTV-HPT, containing an expression cassette for the hpt I gene which confers hygromycin resistance. After a period of selection and growth on nutrient agar, several independent transformants were identified, by immunoblotting, that expressed the chimeric ICAM-1 molecule, the J chain and SC in several states of assembly.

FIG. 3 illustrates the expression of the chimeric ICAM-1 molecule in independently transformed tobacco calli. FIG. 3A shows immunoblots of non-reducing SDS-polyacrylamide gels on which samples containing different transformed tobacco calli (C) and aqueous extracts (Aq) were run and probed for the presence of human ICAM. The solubility of the immunoadhesin assured us that extraction could be easily performed, and the similarity of signals leads us to believe in the reproducibility of expression. FIG. 3B shows immunoblots of nonreducing SDS-polyacrylamide gels containing various fractions of partially purified immunoadhesin from callus Rhi107-11. The blots were probed with antibodies against human ICAM (~ICAM), human IgA heavy chain (~α), human secretory component (~SC) and human J chain (~J). Secondary, enzyme-conjugated antibodies were employed as necessary to label immuno-positive bands with alkaline phosphatase. The specificity of immuno-blotting was further verified by a failure to detect immuno-positive bands in extracts of non-expressing calli (not shown). The expected MW for a dimerized chimeric ICAM-1 molecule, without glycosylation, is 173,318; this form is likely present in the band migrating just below the 250kD marker since it is immuno-positive for ICAM-1 and heavy-chain. This band is also immuno-positive for SC (total expected MW of ~248 kD) but not for J chain which is somewhat unexpected given the canonical pathway for SIgA

assembly, which involves 2 cell types (in mammalian) and requires the presence of J chain prior to assembly of SC. A tetrameric immunoadhesin, containing a single molecule of J chain and a single molecule of SC, has an expected MW of ~440 kD, prior to glycosylation. Several species with molecular weights well in excess of 200 kD, immuno-positive with all four probes, are readily apparent.

Bombardment with DNA-coated microprojectiles is used to produce stable transformants in both plants and animals (reviewed by Sanford *et al.*, *Meth. Enz.* 217:483-509,1993). Particle-mediated transformation with the vectors encoding the immunoadhesin of the present invention was performed using the PDS-1000/He particle acceleration device, manufactured by Bio-Rad. The PDS-1000/He particle acceleration device system uses Helium pressure to accelerate DNA-coated microparticles toward target cells. The physical nature of the technique makes it extremely versatile and easy to use. We have successfully transformed tobacco with all four components of a secretory IgA simultaneously.

The basic biolistic procedure was performed as follows: A stock suspension of microprojectiles was prepared by mixing 60 mg of particles in 1 ml of absolute ethanol. This suspension was vortexed and 25-50 µl was removed and added to a sterile microcentrifuge tube. After microcentrifuging for 30 seconds the ethanol was removed and the pellet resuspended in 1 ml sterile water and centrifuged for 5 minutes. The water was then removed and the pellet resuspended in 25-50 µl of DNA solution containing a mixture of plasmid DNAs, usually, but not always in equimolar amounts. The amount of plasmid added varied between 0.5 ng and 1 µg per preparation. The following were added sequentially: 220 µl of sterile water, 250 µl of 2.5M CaCl2, and 50 µl of 0.1M spermidine. This mixture was vortexed for at least 10 min and then centrifuged for 5 min. The supernatant was removed and the DNA/microprojectile precipitated in 600 µl of absolute ethanol, mixed and centrifuged 1min. The ethanol was removed and the pellet resuspended in 36 µl of ethanol. Ten µl of the suspension was applied as evenly as possible onto the center of a macrocarrier sheet made of Kapton (DuPont) and the ethanol was evaporated. The macrocarrier sheet and a rupture disk were placed in the unit. A petri dish containing surface-sterilized tobacco

leaves was placed below the stopping screen. The chamber was evacuated to 28-29mm Hg and the target was bombarded once. The protocol has been optimized for tobacco, but is optimized for other plants as well by varying parameters such as He pressure, quantity of coated particles, distance between the macrocarrier and the stopping screen and flying distance from the stopping screen to the tissue.

Expression cassettes for chimeric ICAM-1 molecules were also assembled in binary vectors for use in *Agrobacterium*-mediated transformation. An *Agrobacterium* binary vector designed for expression of both human J chain and human secretory component, as well as kanamycin resistance, was introduced into *A.* tumefaciens strain LBA4404. The chimeric ICAM/IgA molecule in another binary vector was also used to transform LBA4404. Overnight cultures of both strains were used for simultaneous "co-cultivation" with leaf pieces of tobacco, according to a standard protocol (Horsch *et al.*, *Science* 227:1229-1231, 1985).

A standard protocol for regeneration of both bombarded and Agrobacterium-transformed tobacco leaf disks was used (Horsch *et al.*, *Science* 227:1229-1231, 1985). Because transformed plants, regenerated from bombarded tissue, frequently undergo gene-silencing upon maturation, transgenic tobacco plants were prepared via *Agrobacterium*-mediated transformation, which gives a higher yield of expressing, mature plants.

3. Purification of Assembled Immunoadhesin

The immunoadhesin expressed according to Examples 3 was purified. Calli were grown in large amounts to facilitate the development of extraction procedures. A partial purification schedule provided a stable concentrate, available in a variety of buffer conditions, for investigation of subsequent chromatographic techniques for the further purification of the immunoadhesin (See **FIG. 3**). Calli were extracted in a juicer, which crushes tissue between two stainless-steel gears, while bathed in a buffer containing sodium citrate (0.6 M, pH 7.4) and urea (final concentration of 2 M). The juice (~1 ml/g fresh weight) was precipitated, after coarse filtration through cheesecloth, with 0.67 volumes of saturated ammonium sulfate. A

green pellet was collected after centrifugation and thoroughly extracted, in a small volume of 50 mM sodium citrate (pH 6.6), with a Dounce homogenizer. After additional centrifugation, a clear brown supernatant was collected and partially purified, during buffer exchange in a de-salting mode, by passage through a Sephadex G-100 column. The desalting/buffer exchange step has allowed preparation of a partially purified concentrate (~0.2 ml/ g fresh weight callus) in a desirable buffer; the G-100 column was eluted with 0.25 X phosphate buffered saline. This eluate appeared to be stable for at least 10 days at 2-8°C.

4. **The Immunoadhesin Inhibits Human Rhinovirus Infectivity**

The infectivity of cells by human rhinovirus was demonstrated to be inhibited by the immunoadhesin prepared according to Example 3. Callus extract prepared according to Example 3 successfully competed for binding of an anti-ICAM monoclonal antibody to soluble ICAM-1. FIG. 4 shows the data from an enzyme-linked immunosorbent assay (ELISA). For the assay, 96-well plates were coated with 0.25 μg soluble ICAM-1/ml. The squares represent the increasing concentrations of sICAM and the circles represent the increasing amounts of callus extract (sterile filtered fraction from G-100) used to compete with the adhered ICAM for a constant amount of a mouse (anti-human ICAM) antibody. After washing the wells, adherent mouse antibody was detected with an anti-mouse antibody conjugated to horseradish peroxidase. Adherent enzyme activity was measured at 490 nm, with ortho-phenylene diamine as a substrate. The data (squares, sICAM; circles, Extract) are well described by sigmoids of the form OD490 = y =y_0 + a/[1+e^-{(x-x_0)/b}], where a = y max, y_0 = y min, b = the slope of the rapidly changing portion of the curve and x_0 = the value of x at the 50% response level. Relative to soluble ICAM-1, the immunoadhesin extract tested here contains the equivalent of ~250 μg ICAM/ml; this is an overestimate due to expected avidity effects of the dimeric and tetrameric assemblies of the ICAM-1-heavy-chain fusions. Thus, this ELISA demonstrated that the immunoadhesin competes with soluble ICAM-1 for binding to an anti-ICAM mAb.

The competitive ELISA allows for quantitative assessment of the recovery of activity by comparing the normalized amounts of various fractions required to give a 50% response.

Upon purification, the titer of a immunoadhesin preparation may be expressed as a reciprocal dilution, or the number of milliliters to which a milligram of immunoadhesin must be diluted in order to give a 50 % response. This ELISA will facilitate the development of a purification process for the immunoadhesin.

A cytopathic effect assay (CPE) demonstrated the specific ability of the partially purified immunoadhesin to inhibit the infectivity of human cells by human rhinovirus (**FIG. 5**). Rhinovirus serotype HRV-39 was pre-incubated with human ICAM-1, an ICAM/IgA fusion (gift of Dr. Tim Springer), or with extracts from calli either expressing our ICAM-1/SIgA immunoadhesin or another, different, antibody before plating each of the mixtures with HeLa S3 cells at 33°C. After 3 days, viable cells were fixed and stained with a methanolic solution of Crystal Violet; the optical density at 570 nm provides a proportional measure of cell viability.

Two extracts derived from Rhi107-11, containing the immunoadhesin, clearly inhibited the virus' ability to infect and kill HeLa S3 cells (**FIG. 5A**, right side-up and upside-down triangles). Because the extracts were only partially purified, we also assayed a similarly prepared extract that contained a human IgA2m(2) directed against Doxorubicin, a chemotherapeutic agent. That extract, containing a similar immunoglobulin with an unrelated binding specificity, was unable to inhibit the infectivity of the rhinovirus and demonstrates that expression of the ICAM-1-heavy-chain fusion confers specificity to the inhibition. The CPE assay demonstrated, as expected, the differential ability of souluble ICAM-1 and an (IC1-5/IgA; Martin, *et al.*, 1993) to inhibit viral infectivity (**FIG. 5B**). The insert in Figure 5B is the scale expansion in the range of the IC50 for soluble ICAM-1 (1.35 μg/ml) and for the ICI-5/IgA (0.12 μg/ml; 11.3 fold less).

5. Production and Purification of Immunoadhesins for Clinical and Toxicological Studies

Production of sufficient immunoadhesin for the proposed clinical and toxicological needs is performed by making transgenic tobacco plants. The transgenic plants which express

the immunoadhesin (without an ER retention signal) are generated by *Agrobacterium*-mediated transformation. The absence of an ER retention signal is anticipated to enhance assembly since the nascent SIgA is processed through the entire Golgi apparatus, including, in particular, the trans-Golgi, where SC is covalently linked to dIgA as suggested by pulse-chase experiments (Chintalacharuvu & Morrison, *Immunotechnology* 4:165-174, 1999). Because *Agrobacterium*-mediated transformation is much more likely to generate plants with consistent levels of transgene expression, it is likely that progeny of these plants will be used for the production of clinical grade immunoadhesin.

In order to maximize expression levels, and create a true-breeding line, it is desirable to create homozygous plants. The highest producing plants (generation T_0) can self-fertilize in the greenhouse before seed is collected. One quarter of the T_1 plants are expected to be homozygous. These are grown in the greenhouse and seed samples from several plants are separately germinated on medium containing kanamycin. All the progeny (T_2) from homozygous positive plants are expected to be green. Some of the progeny of heterozygous plants are expected to be white or yellowish. Homozygosity is confirmed by back-crossing to wild-type and immunoblotting extracts of the progeny.

Harvesting and processing may be continuously meshed during a production campaign, especially since multiple harvests may be obtained from a single planting, i.e. plants cut to soil level for one harvest are regrown for subsequent harvests. In developing a sense of scale for the production of immunoadhesin it is necessary to decide on the required amount of finished immunoadhesin, account for expression levels (mg immunoadhesin present/ kg fresh weight tobacco), know the growth rate of the plants and the expected weight of the average plant, and the overall yield of the purification schedule (set at 20%). Setting the overall need at 3 g of finished immunoadhesin requires preparing for 4 harvests, each with an expected yield of 1 g of finished immunoadhesin.

Given these targets and parameters, the necessary number of plants and hence the space requirements for plant growth is determined. **FIG. 6** shows an evaluation of the

production necessities for making 1 gram of finished Immunoadhesin. In this diagram, the number of plants needed for 1 g of immunoadhesin, at 20% yield, at expected levels of expression and plant weight is illustrated. At different levels of immunoadhesin expression (mg/kg fresh weight) and overall recovery (set at 20%), the weight of each plant, and so the total number of plants, may be determined for a specified production target (1 g/harvest) within a window (dotted square) of reasonable possibilities. The number of required plants decreases, inversely, with the number of specified growth and re-growth periods. The expected biomass production, a function of time and growth conditions, influences the time to harvest and the time between harvests. These growth periods can be adjusted to the realities of the purification schedule by staggering planting and harvesting dates. From our experience, production requires x number of plants. For example, 1 g of finished immunoadhesin from plants with a reasonable expression level, of 100 mg of immunoadhesin/kg fresh weight, require 250 plants when harvested at a weight of 200 g/plant (~80 days post germination). At this scale, these plants require about 10 m^2 of growing space and are harvested twice over 150 days.

Processing 50+ kg of biomass at a time requires several moderately large-scale operations which all have counter-parts in the food-processing industry. These include bulk materials handling, size reduction, juicing and filtration. A Vincent Press and a Durco filtration system are used to efficiently process these quantities. The juicing step employs a proven and simple buffer of sodium citrate and urea. These components buffer the extract, help prevent the oxidation of phenolics and their association with proteins (Gegenheimer, *Methods in Enzymology* 182:174-193, 1990; Loomis, *Methods in Enzymology*, 31:528-544, 1974; Van Sumere, *et al., The Chemistry and Biochemistry of Plant Proteins,* 1975.) and ensure the solubility of the immunoadhesin during a subsequent acid precipitation.

Filtration of acid-insoluble lipid and protein (~90% of the total) is followed by tangential flow ultrafiltration to concentrate the immunoadhesin and to remove small proteins, especially phenolics. Diafiltration enhances the removal of small molecules and exchanges the buffer in preparation for short-term storage and subsequent chromatography. Either

SP-Sepharose (binding at pH 5.0 or below) or Q-Sepharose (binding at pH 5.5 or above) are among the ion-exchanges that can be used for filtering immunoadhesin. They are readily available, scalable, robust and have high capacities. In particular, they are available for expanded-bed formats, which reduce the stringency of prior filtration steps. Cation-exchange chromatography, which can be more selective than anion-exchange chromatography, is used first. The immunoadhesin is purified from the several species of protein potentially present, to the point where at least 95% of the protein is in the form of ICAM-1/IgA, ICAM-1/dIgA or ICAM-1/SIgA, as the presence of di- and tetra-valent ICAM-1 domains are critical for potent anti-viral activity. Purified immunoadhesin is then tested for acceptable levels of endotoxin, alkaloids such as nicotine and for bio-burden. In addition, potency levels (defined by ELISA and CPE assays), protein concentration, pH and appearance are monitored. Subsequently, the stability of the clinical lots of immunoadhesin is determined, to ensure that patients receive fully potent immunoadhesin. Even partially purified extracts have been found to be stable for 10 days when refrigerated. The titer and potency of clinically formulated immunoadhesin (in phosphate-buffered saline), when stored at –20°C, 2-8°C, and at 37°C, over a period of 3 to 6 months, is also tested.

6. **The Immunoadhesins Have Plant-Specific Glycosylation**

The immunoadhesins produced are analyzed to determine the pattern of glycosylation present. Cabanes-Macheteau *et al.*,*Glycobiology* 9(4):365-372 (1999), demonstrated the presence of several glycosyl moieties, typical of plants, on a plant-expressed antibody construct. Their methods are used to demonstrate that the immunoadhesins produced according to Example 1, 2 and 3 have a plant-specific glycosylation pattern. We anticipate that this diversity will also be a source of variability for immunoadhesin. Since crude extracts have been shown to have anti-viral activity *in vitro* (data not shown), glycosylation, as such, does not appear to affect potency. N-linked glycosylation (**FIG. 2** shows that there are fifteen potential sites on the chimeric ICAM-1 molecule alone) probably contributes to the diversity of bands seen in immuno-blots. Immunoadhesin preparations are digested with N-Glycosidase A, before blotting, showing that the difference in banding patterns collapse

into fewer, discrete bands. In this way, glycoforms are initially characterized with reducing and non-reducing polyacrylamide gels. In addition, digested and mock-digested fractions are tested in the CPE assay and competition ELISA, demonstrating the effect of N-linked glycosylation on potency and titer *in vitro*.

7. The Immunoadhesin Inactivates Human Rhinovirus

The immunoadhesin prepared according to Examples 1, 2 and 3 is assayed for its ability and to inactivate HRV by binding to the virus, blocking virus entry, and inducing the formation of empty virus capsids. To measure binding of the immunoadhesin to HRV, the immunoadhesin is incubated with [^{3}H]leucine-labeled HRV-39 for 30 min and then added to HeLa cells for 1 hr. After washing, cells and bound virus are detached with Triton X-100 and [^{3}H] measured in a scintillation counter.

Inactivation of HRV-39 by incubation with the immunoadhesin is compared with HRV inactivation by sICAM-1. HRV-39 is not directly inactivated to a significant extent (<0.5 $\log_{10}$ reduction in infectivity) by incubation with monomeric sICAM-1, while incubation with IC1-5D/IgA reduced infectivity approximately 1.0 $\log_{10}$ (Arruda, *et al.*, *Antimicrob. Agents Chemother.* 36:1186-1191, 1992; Crump, *et al., Antimicrob. Agents Chemother.* 38:1425-7, 1994). In order to test the ability of the immunoadhesin to inactivate HRV-39, 10^6 50% tissue culture infective doses ($TCID_{50}$) of HRV-39 are incubated in medium containing a concentration of sICAM-1 or immunoadhesin equal to ten times the IC_{50} of each molecule for that virus, or in plain medium, for 1 hr at 33°C on a rocker platform. Each virus-immunoadhesin or virus-medium mixture are then diluted serially in ten-fold dilutions, and the titer determined on HeLa cells in 96-well plates.

The effect of the immunoadhesin on HRV attachment to host cells is tested by inoculating HeLa cells with HRV-39 at a MOI of 0.3 in the presence or absence of the immunoadhesin. Absorbance proceeds for one hour at 4°C, the cells are washed, and media is replaced plus or minus the immunoadhesin. Cells are incubated for ten hours at 33°C (to

allow one round of replication), and virus are harvested by freeze/thawing the cells. The virus is titered on HeLa cells.

ICAM-IgA (IC1-5D/IgA) is more efficient than sICAM-1 at inducing conformational changes in HRV, leading to the formation of empty, non-infectious viral particles (Martin, *et al. J. Virol.* 67:3561-8, 1993). To examine the ability of the immunoadhesin produced according to Examples 1, 2 and 3 to induce conformational changes in HRV, causing release of viral RNA, purified immunoadhesin is incubated with [^{3}H]leucine-labeled HRV-39 for 30 min and then the virus is overlayed onto a 5 to 30% sucrose gradient. Following centrifugation for 90 min at 40,000 rpm, fractions are collected, [^{3}H] measured, and fractions assessed for infectivity. (Intact HRV sediments at 149S on a sucrose gradient while empty capsids lacking RNA sediments at 75S (Martin, *et al. J. Virol.* 67:3561-8, 1993)). Due to its increased valence, we expect the ICAM/SIgA immunoadhesin is more efficient at inducing empty non-infectious particles than ICAM-IgA.

The inhibitory effect of purified immunoadhesin on a panel of both major and minor (that do not use ICAM-1 as a receptor) HRV serotypes will be examined using the CPE assay. The ability of ICAM-1 to inhibit HRV infection varies among viral isolates. It has been shown (Crump, *et al., Antimicrob. Agents Chemother.* 38:1425-7, 1994) that the EC_{50} for sICAM-1 varies from 0.6 μg/ml to >32 μg/ml when tested on a panel of HRV major receptor serotypes assay using HeLa cells. Our panel includes nine major serotypes (HRV-3, -13, -14, -16, -23, -39, -68, -73, and –80) and the minor receptor serotype HRV-1A.

8. Clinical Studies Demonstrating the Ability of the Immunoadhesin to Reduce Infectivity in Humans: Dose Escalation Tolerance Study

The immunoadhesin of the present invention is tested in two randomized controlled trials to determine the effect of intranasal administration of the immunoadhesin on infection, IL-8 response, and illness in experimental rhinovirus colds. These two studies evaluate the immunoadhesin taken by subjects before or after rhinovirus inoculation. The clinical protocols used here are based on protocols previously used by in evaluation of a recombinant

soluble ICAM-1 molecule for efficacy against rhinovirus infection (Turner, *et al., JAMA* 281:1797-804, 1999).

A. Subjects.

Subjects are recruited from university communities at the University of Virginia, Charlottesville. Subjects are required to be in good health, non-smokers, and between the ages of 18 and 60 years. Subjects are excluded if they have a history of allergic disease or nonallergic rhinitis, abnormal nasal anatomy or mucosa, or a respiratory tract infection in the previous 2 weeks. Pregnant or lactating women or women not taking medically approved birth control are also excluded. In the experimental virus challenge study (Phase I/II, see below), subjects are required to be susceptible to the study virus as evidenced by a serum neutralizing antibody titer of 1:4 or less to the virus, determined within 90 days of the start of the trial.

B. Study Medication.

The immunoadhesin of the present invention is formulated as a phosphate-buffered saline (PBS) spray solution containing 2.6 mg/ml. The placebo consists of PBS and is identical in appearance to the active preparation. The solutions are administered using a medication bottle equipped with a metered nasal spray pump. The pump delivers 70 µl of solution containing 183 µg of the immunoadhesin with each spray. The medication is administered as two sprays per nostril, six times daily (at 3-hour intervals) for a total daily dose of 4.4 mg. This is the same dose, in mg protein/day, as was used for soluble ICAM-1 in the tremacamra study infection (Turner, *et al., JAMA* 281:1797-804, 1999). A mole of the immunoadhesin has about twice the mass as a mole of sICAM-1. However, given the differences in *in vitro* activity between sICAM-1 and ICAM/IgA fusions, the immunoadhesin is many fold more effective on a molar basis than sICAM-1. Thus, this amount is a conservative calculation of what is necessary. This amount is used, except in the event that the dose escalation study reveals problems at this dose.

C. Study Design

Single ascending dose and multiple ascending dose studies are used to evaluate the safety of the immunoadhesin. In each case, three subjects are evaluated at each dosage level, two receiving the immunoadhesin and one receiving placebo. In the single ascending dose study, four dosage levels are evaluated. The lowest individual dose is half the anticipated dose to be used in the challenge study, and the highest individual dose is twice the anticipated challenge study dose. The dosage levels are as follows: one spray in each nostril (366 μg total), two sprays in each nostril (732 μg total), three sprays in each nostril (1098 μg total), four sprays in each nostril (1464 μg total).

The same dosage levels are used in the multiple ascending dose study. Subjects receive doses every three hours (six times per day) for five days. In both studies subjects are evaluated at each dosage level, staggering the start of each subsequent level until it is clear that there is no acute toxicity at the previous level. All subjects return for a single dose 21 days after the first dose, and then for a follow-up at six weeks (for determination of serum antibody against the immunoadhesin).

A separate group of twelve subjects is given one dose of two sprays in each nostril (732 μg total), and nasal lavage is done at 1, 2, 4, 8 and 16 hours (two subjects at each time point). Washings are assayed at Panorama Research by ELISA for the immunoadhesin in order to calculate its *in vivo* half-life. The total amount of the immunoadhesin to be used in the dose escalation and half-life determination studies (on a total of 28 subjects) will be approximately 270 mg.

D. Safety Evaluations.

In addition to routine adverse event recording, the safety of the immunoadhesin is assessed in three ways. First, prior to the first dose and after the last dose the investigators perform a visual examination of the nasal mucosa, in particular looking for signs of irritation or inflammation. Any visible changes are noted. Second, standard blood safety evaluations are done on samples collected prior to treatment and after the last dose on study days 1, 4, and

8 (and 21 in the multiple ascending dose study). Third, serum samples are saved, frozen, and used to determine if the immunoadhesin is able to pass through the nasal mucosa into the blood. This is accomplished in two ways. First, the presence the immunoadhesin in serum samples is measured by ELISA. In this assay, anti-human IgA antibodies adsorbed to microtiter plates capture any the immunoadhesin in the serum, which are detected by an anti-ICAM antibody. The sensitivity of the assay is determined using normal human serum samples spiked with known concentrations of the immunoadhesin. Alternatively, anti-ICAM antibodies can be adsorbed to plates to capture the immunoadhesin in the serum, that would be detected by anti-IgA. Second, the presence of an immune response to the immunoadhesin is assayed with an ELISA method that uses the immunoadhesin adsorbed to microtiter plates. Any anti-immunoadhesin antibodies in the serum bind, and are detected with anti-human IgG or anti-human IgM. Pre-treatment and post-treatment serum samples are compared, and any change in titer is considered evidence of uptake of the immunoadhesin. If there is any positive evidence of anti-immunoadhesin antibodies, additional assays will be done to distinguish between anti-ICAM-1 and anti-IgA activity.

Patients are screened for the development of an allergic reaction to the immunoadhesin. (In previous studies, there were no episodes of adverse reactions with soluble ICAM applied topically in the nose or plantibodies applied topically in the oral cavity.) Individuals exhibiting symptoms of nasal allergy are tested for anti-immunoadhesin-specific IgE antibodies in nasal lavage fluids using a sensitive two-step ELISA (R & D Systems).

E. Statistical Analysis.

The sample size for these studies is based on previous studies using the rhinovirus challenge model. The sample size planned for the protection studies should be adequate to detect a reduction in the incidence of clinical colds from 75% in the placebo groups to 25% in the active treatment groups at 1-sided levels of $\alpha = .05$ and $1-\beta = .80$. In addition, the sample size should be adequate to detect a change in the total symptom score of 5 units assuming an SD of 5.8 units.

9. Clinical Studies Demonstrating the Ability of the Immunoadhesin to Reduce Infectivity in Humans: Challenge Studies

Challenge studies are used to demonstrate that treatment with the immunoadhesin of the present invention protect against clinical colds or reduce cold symptoms after viral challenge.

A. Challenge Virus.

The challenge virus used for this study is rhinovirus 39 (HRV-39). Rhinovirus type 39 is a major group of rhinovirus that requires ICAM-1 for attachment to cells. The challenge virus pool is safety-tested according to consensus guidelines (Gwaltney, *et al.*, *Prog. Med. Virol.* 39:256-263, 1992). All subjects are inoculated with approximately 200 median tissue culture infective dose ($TCID_{50}$). The virus are administered as drops in two inocula of 250 µl per nostril given approximately 15 minutes apart while the subjects are supine.

TABLE 1

				Pre-inoculation study timetable					
				Day					
	0	1	2	3	4	5	6	7 - 14	21
Medications		6 doses	6 doses	6 doses	6 doses	6 doses			
Inoculation		hour 4							
Symptom scores		m/e	m/e	m/e	m/e	m/e	m/e	e	
Nasal lavage		m	m	m	m	m	m		
Serum sample	X								X
				Post-inoculation study timetable					
				Day					
	0	1	2	3	4	5	6	7 - 14	21
Medications		6 doses	6 doses	6 doses	6 doses	6 doses			
Inoculation	hour 0								
Symptom scores		m/e	m/e	m/e	m/e	m/e	m/e	e	
Nasal lavage		m	m	m	m	m	m		
Serum sample	X								X
Note: In both studies on days 1-5, doses are given at hours 0, 3, 6, 9, 12, and 15									
m = morning									
e = evening									

B. Study Design.

Two randomized rhinovirus challenge studies are performed (see Table 1). The same formulation of the immunoadhesin of the present invention is evaluated in pre-inoculation and post-inoculation studies. In both studies, medication is administered as six doses each day for five days. Subjects are randomly assigned to receive either the immunoadhesin or matching placebo at the time of enrollment into each study. The study is blinded and all clinical trial personnel, subjects, and employees of Panorama Research remain blinded until all data are collected.

In the pre-inoculation study, medications are started four hours (two doses) prior to viral challenge. The virus challenge is administered one hour after the second dose of the

immunoadhesin (or placebo) and the four remaining doses of study medication for the first day are given as scheduled. In this study eighteen subjects receive the active treatment and eighteen subjects receive placebo.

In the post-inoculation study, medications begin 24 hours after virus challenge. This timepoint was chosen because it has been used in other studies of protection from virus challenge, and because cold symptoms are clearly present (Harris & Gwaltney, *Clin. Infect. Dis.* 23:1287-90, 1996). Virus challenge in this study is administered in the morning of study day 0 approximately 24 hours prior to the first dose of study medication on the morning of study day 1. In this study, 36 subjects receive the active treatment and 18 subjects receive placebo.

Subjects are isolated in individual hotel rooms from study day 0 (the day of virus challenge) to study day 6. On each of these days a symptom score and a nasal lavage for virus isolation are done in the morning prior to the first dose of medication and a second symptom score is done each evening. On study day 6, subjects are released from isolation but continue to record symptom scores each evening through day 14. The subjects return to the study site on study day 21, when a final serum sample for detection of anti-immunoadhesin antibodies will be collected. The total amount of immunoadhesin to be used in the two virus challenge studies (on a total of 54 subjects) is approximately 1200 mg.

C. Viral Isolation.

Virus shedding is detected by virus isolation in cell culture. Nasal wash specimens are collected by instillation of 5 ml of 0.9% saline into each nostril. This wash is then expelled into a plastic cup and kept chilled for one to two hours until it is processed for viral cultures. Immunoadhesin is removed from the specimens by treatment with anti-ICAM-1 antibody adsorbed to an agarose support (Affi-Gel 10, Bio-Rad Laboratories, Hercules, CA). A portion of each processed specimen is stored at –80 °C, and another portion is inoculated into two tubes of HeLa-1 cells, a HeLa cell line enriched for the production of ICAM-1 Arruda, *et al.*, *J. Clin. Microb.* 34:1277-1279, 1996). Rhinovirus are identified by the development of

typical cytopathic effect. Subjects with a positive viral culture on any of the postchallenge study days are considered infected. Viral titers in the specimens stored at –80 °C are determined by culturing serial ten-fold dilutions in microtiter plates of HeLa-1 cells.

Antibody to the challenge virus are detected by serum neutralizing titers done using standard methods Gwaltney, *et al, Diagnostic Procedures for Viral Rickettsial and Chlamydial Infections*, p. 579-614, American Public Health Association). Serum specimens for antibody testing are collected during screening, immediately prior to virus challenge (acute), and again 21 days later (convalescent). Subjects with at least a four-fold rise in antibody titer to the challenge virus when the convalescent serum sample is compared with the acute serum sample are considered infected.

D. Evaluation of Illness Severity

Illness severity is assessed as previously described (Turner, *et al., JAMA* 281:1797-804, 1999). Symptom scores are recorded prior to virus challenge (baseline) and twice each day at approximately twelve-hour intervals for the next 6 days. On study days 7 through 14 each subject records his/her symptom score once per day in the evening. At each evaluation, subjects are asked to judge the maximum severity of the following eight symptoms in the interval since the last symptom evaluation: sneezing, rhinorrhea, nasal obstruction, sore throat, cough, headache, malaise, and chilliness. Each symptom is assigned a severity score of 0 to 3 corresponding to a report of symptom severity of absent, mild, moderate, or severe. If symptoms are present at baseline, the baseline symptom score will be subtracted from the reported symptom score. The higher of the two daily evaluations are taken as the daily symptom score for each symptom. The daily symptom scores for the eight individual symptoms are summed to yield the total daily symptom score. The total daily symptom scores for the first 5 days after virus challenge (study days 1-5) are summed and on the evening of study day 5, all subjects are asked, "Do you feel you have had a cold?" Subjects who had a total symptom score of at least 6 and either at least three days of

rhinorrhea or the subjective impression that they had a cold are defined as having a clinical cold.

The weight of expelled nasal secretions is determined on days 1-7 by providing all subjects with packets of preweighed nasal tissues. After the tissues are used they are stored in an airtight plastic bag. Each morning the used tissues, together with any unused tissues from the original packet, are collected and weighed.

E. IL-8 Assay.

Recent studies have suggested that the host inflammatory response, particularly interleukin 8 (IL-8), may play a role in the pathogenesis of common cold symptoms due to rhinovirus infection. Concentrations of IL-8 in nasal lavage are determined with a commercially available ELISA (R&D Systems, Minneapolis, Minn) as previously described (Turner, *et al., JAMA* 281:1797-804, 1999).

F. Safety Evaluations.

The same evaluations are done in the challenge study as in the dose escalation study described in Example 8.

G. Statistical Analysis.

Statistical analysis is performed similarly as to that described for the dose escalation study described in Example 8.

Claims

1. An immunoadhesin comprising a chimeric ICAM-1 molecule, said chimeric ICAM-1 molecule comprising:

a rhinovirus receptor protein linked to at least a portion of an immunoglobulin heavy chain; and

J chain and secretory component associated with said chimeric ICAM-1 molecule.

2. The immunoadhesin of claim 1 wherein said rhinovirus receptor protein is comprised of any combination of extracellular domains 1, 2, 3, 4 and 5 of ICAM-1.

3. The immunoadhesin of claim 1 wherein said immunoglobulin is selected from the group of IgA, IgA_1, IgA_2, IgM, and chimeric immunoglobulin heavy chains.

4. The immunoadhesin of claim 1 further comprising at least one additional chimeric ICAM-1 molecule.

5. The immunoadhesin of claim 1 wherein said rhinovirus receptor protein is comprised of any combination of extracellular domains 1, 2, 3, 4 and 5 of ICAM-1; and said immunoglobulin heavy chain comprises at least a portion of an IgA_2 heavy chain.

6. The immunoadhesin of claim 1 expressed in transgenic plants.

7. The immunoadhesin of claim 1 expressed in monocotyledonous plants.

8. The immunoadhesin of claim 1 expressed in dicotyledonous plants.

9. The immunoadhesin of claim 1 wherein all proteins are human.

10. The immunoadhesin of claim 1 expressed in heterologous cells derived from plants, vertebrates, or invertebrates.

11. The immunoadhesin of claim 1 expressed in mammalian cells.

12. The immunoadhesin of claim 1 expressed in hairy root cultures.

13. The immunoadhesin of claim 1 expressed in plant cells in tissue culture.

14. An immunoadhesin comprising a chimeric ICAM-1 molecule, said chimeric ICAM-1 molecule comprising: a rhinovirus receptor protein linked to at least a portion of an immunoglobulin heavy chain, wherein immunoadhesin has plant-specific glycosylation.

15. The immunoadhesin of claim 14 wherein said immunoadhesin further comprises a J chain and secretory component associated with said chimeric ICAM-1 molecule.

16. The immunoadhesin of claim 14 wherein said rhinovirus receptor protein is comprised of any combination of extracellular domains 1, 2, 3, 4 and 5 of ICAM-1.

17. The immunoadhesin of claim 14 wherein said immunoglobulin heavy chain is selected from the group of IgA, IgA_1, IgA_2, IgG_1, IgG_2, IgG_3, IgG_4, IgM, IgD, IgE, and a chimeric immunoglobulin heavy chain.

18. The immunoadhesin of claim 14 further comprising at least one additional chimeric ICAM-1 molecule.

19. The immunoadhesin of claim 14 wherein said rhinovirus receptor protein is comprised of any combination of extracellular domains 1, 2, 3, 4 and 5 of ICAM-1; and said immunoglobulin heavy chain comprises at least a portion of an IgA_2 heavy chain.

20. The immunoadhesin of claim 14 wherein all proteins are human.

21. The immunoadhesin of claim 14 expressed in heterologous cells derived from plants.

22. The immunoadhesin of claim 14 expressed in hairy root cultures.

23. The immunoadhesin of claim 14 expressed in plant cells in tissue culture.

24. The immunoadhesin of claim 14 expressed in plants.

25. The immunoadhesin of claim 14 expressed in monocotyledonous plants.

26. The immunoadhesin of claim 14 expressed in dicotyledonous plants.

27. A composition comprising an immunoadhesin and plant material, wherein said immunoadhesin comprises a chimeric ICAM-1 molecule, said chimeric ICAM-1 molecule comprising a rhinovirus receptor protein linked to at least a portion of an immunoglobulin heavy chain.

28. The composition of claim 27 further comprising a J chain and secretory component associated with said chimeric ICAM-1 molecule.

29. A composition of claim 27 wherein said rhinovirus receptor protein is comprised of any combination of extracellular domains 1, 2, 3, 4 and 5 of ICAM-1; and said immunoadhesin has plant-specific glycosylation.

30. A composition of claim 27 wherein said immunoglobulin is selected from the group of IgA, IgA_1, IgA_2, IgG_1, IgG_2, IgG_3, IgG_4, IgM, IgD, IgE, and a chimeric immunoglobulin heavy chain.

31. A composition of claim 27 further comprising at least one additional chimeric ICAM-1 molecule.

32. A composition of claim 27 wherein said rhinovirus receptor protein is comprised of any combination of extracellular domains 1, 2, 3, 4 and 5 of ICAM-1; and said immunoglobulin heavy chain is an IgA_2 heavy chain

33. A method for reducing the infection by human rhinovirus of host cells susceptible to infection by human rhinovirus, said method comprising:

contacting the virus with an immunoadhesin of claim 1, 14, or 27, and wherein

said immunoadhesin binds to human rhinovirus and reduces infectivity thereof.

34. A method for reducing the initiation or spread of the common cold due to human rhinovirus, said method comprising:

contacting the virus with an immunoadhesin of claim 1, 14 or 27, and wherein

said immunoadhesin binds to human rhinovirus and reduces infectivity thereof.

35. A method for the treatment or prevention of human rhinovirus infection in a human subject, said method comprising:

administering to said subject an effective amount of an immunoadhesin of claim 1, 14 or 27, and wherein said immunoadhesin reduces human rhinovirus infectivity thereof.

36. A method for the treatment or prevention of human rhinovirus infection in a subject, said method comprising:

intranasally administering to said subject an effective amount of an immunoadhesin of claim 1, 14 or 27, and wherein said immunoadhesin reduces human rhinovirus infectivity thereof.

37. A method for the treatment or prevention of human rhinovirus infection in a subject, said method comprising: administering through the oral cavity to said subject an effective amount of an immunoadhesin of claim 1, 14 or 27, and wherein said immunoadhesin reduces human rhinovirus infectivity thereof.

38. A pharmaceutical composition comprising an immunoadhesin of claims 1, 14, or 27 in a pharmaceutically acceptable buffer.

39. An expression vector comprising a gene encoding a chimeric ICAM-1 molecule operatively linked to a plant promoter, said chimeric ICAM-1 molecule comprising a rhinovirus receptor protein linked to at least a portion of an immunoglobulin heavy chain.

FIGURE 1

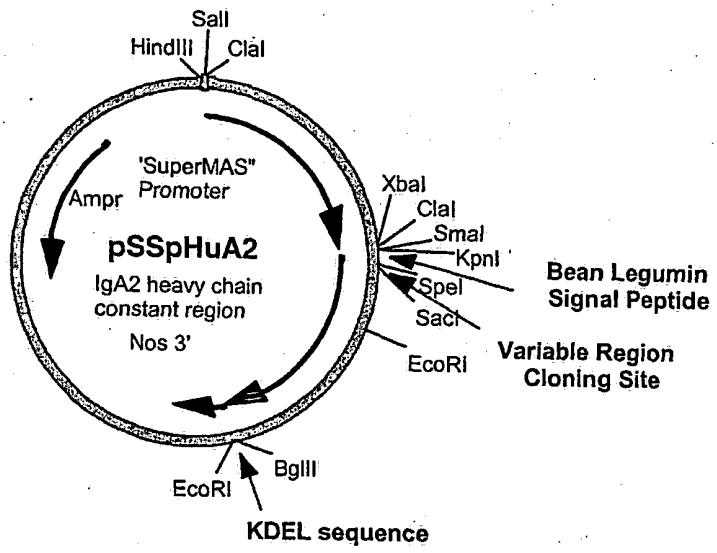
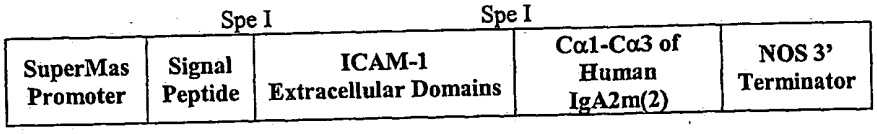

SalI
HindIII
ClaI
'SuperMAS"
Promoter
Ampr
pSSpHuA2
IgA2 heavy chain
constant region
Nos 3'
XbaI
ClaI
SmaI
KpnI
SpeI
SacI
EcoRI
Bean Legumin
Signal Peptide
Variable Region
Cloning Site
EcoRI
BglII
KDEL sequence

Figure 2

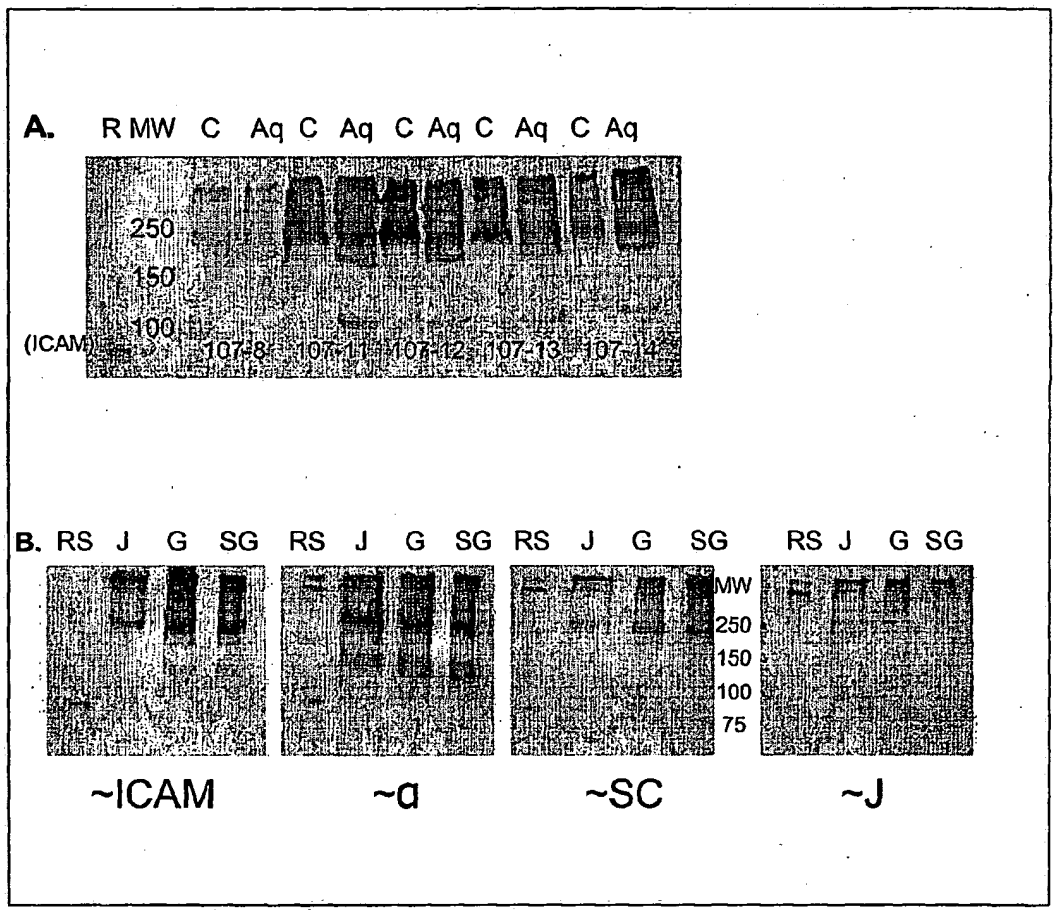
A.



B.

QTSVSPSKVILPRGGSVLVTCSTSCDQPKLLGIETPLPKKELLLPGNNRKVYELSNVQEDSQPMC
YSNCPDGQSTAKTFLTVYWTPERVELAPLPSWQPVGKNLTLRCQVEGGAPRANLTVVLLRGEKELKREPA
VGEPAEVTTTVLVRRDHHGANFSCRTELDLRPQGLELFENTSAPYQLQTFVLPATPPQLVSPRVLEVDTQ
GTVVCSLDGLFPVSEAQVHLALGDQRLNPTVTYGNDSFSAKASVSVTAEDEGTQRLTCAVILGNQSQETL
QTVTIYSFPAPNVILTKPEVSEGTEVTVKCEAHPRAKVTLNGVPAQPLGPRAQLLLKATPEDNGRSFSCS
ATLEVAGQLIHKNQTRELRVLYGPRLDERDCPGNWTWPENSQQTPMCQAWGNPLPELKCLKDGTFPLPIG
ESVTVTRDLEGTYLCRARSTQGEVTREVTVNVTSGSSASPTSPKVFPLSLDSTPQDGNVVVACLVQGFFP
QEPLSVTWSESGQNVTARNFPPSQDASGDLYTTSSQLTLPATQCPDGKSVTCHVKHYTNSSQDVTVPCRV
PPPPPCCHPRLSLHRPALEDLLLGSEANLTCTLTGLRDASGATFTWTPSSGKSAVQGPPERDLCGCYSVS
RVLPGCAQPWNHGETFTCTAAHPELKTPLTANITKSGNTFRPEVHLLPPPSEELALNELVTLTCLARGFS
PKDVLVRWLQGSQELPREKYLTWASRQEPSQGTTTYAVTSILRVAAEDWKKGETFSCMVGHEALPLAFTQ
KTIDRLAGKPTHINVSVVMAEADGTCYRSEKDEL

FIGURE 3


A.
R MW C Aq C Aq C Aq C Aq C Aq
250
150
100
(ICAM)
107-8
107-11
107-12
107-13
107-14
B.
RS J G SG RS J G SG RS J G SG RS J G SG
MW
250
150
100
75
~ICAM
~α
~SC
~J

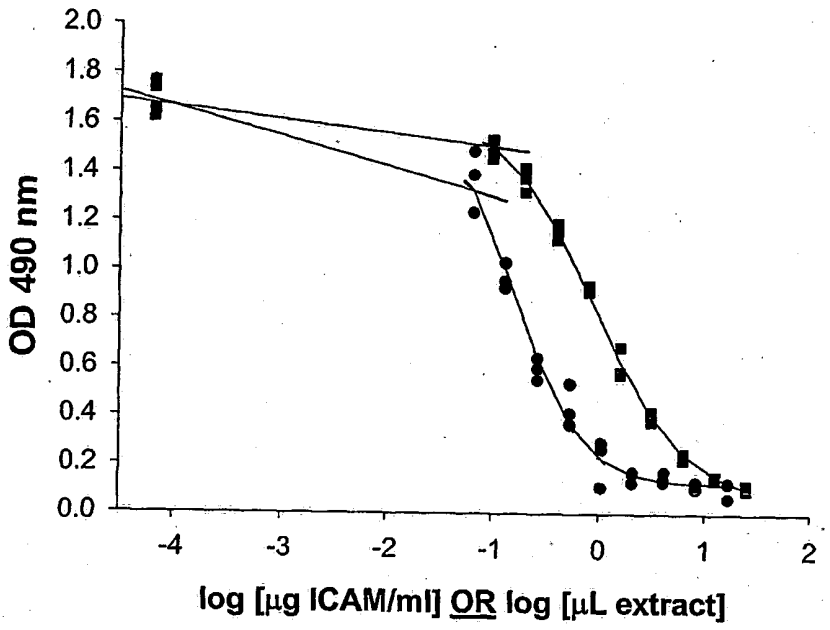

2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
OD 490 nm
-4
-3
-2
-1
0
1
2
log [µg ICAM/ml] OR log [µL extract]

FIGURE 4

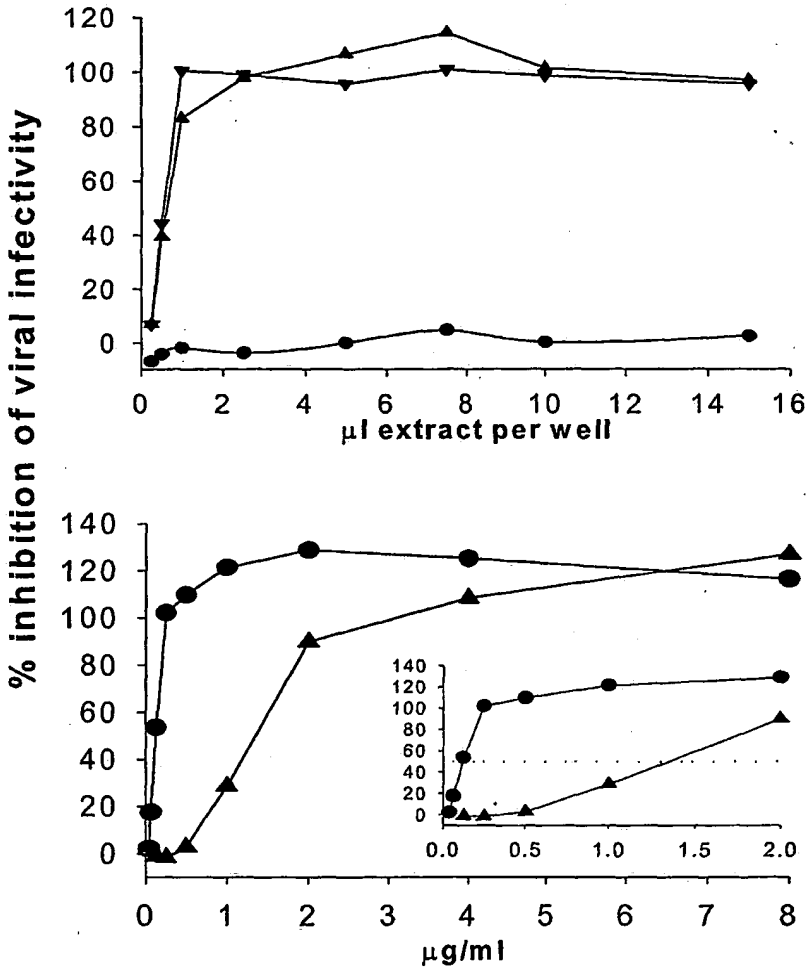

% inhibition of viral infectivity
120
100
80
60
40
20
0
0 2 4 6 8 10 12 14 16
µl extract per well
140
120
100
80
60
40
20
0
0 1 2 3 4 5 6 7 8
µg/ml
140
120
100
80
60
40
20
0
0.0 0.5 1.0 1.5 2.0

FIGURE 5

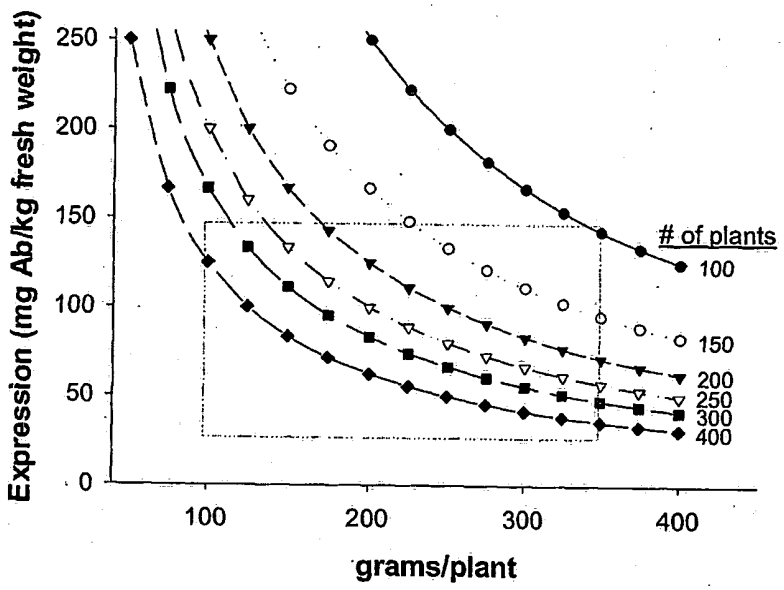

Expression (mg Ab/kg fresh weight)
250
200
150
100
50
0
100
200
300
400
grams/plant
# of plants
100
150
200
250
300
400

FIGURE 6

FIGURE 7 A

I. HUMAN IG ALPHA-1 CHAIN C REGION - HOMO SAPIENS (HUMAN).

AMINO ACID SEQUENCE

>sp|P01876|ALC1_HUMAN IG ALPHA-1 CHAIN C REGION - Homo sapiens (Human).

```
        10         20         30         40         50         60
         |          |          |          |          |          |
ASPTSPKVFP LSLCSTQPDG NVVIACLVQG FFPQEPLSVT WSESGQGVTA RNFPPSQDAS

        70         80         90        100        110        120
         |          |          |          |          |          |
GDLYTTSSQL TLPATQCLAG KSVTCHVKHY TNPSQDVTVP CPVPSTPPTP SPSTPPTPSP

       130        140        150        160        170        180
         |          |          |          |          |          |
SCCHPRLSLH RPALEDLLLG SEANLTCTLT GLRDASGVTF TWTPSSGKSA VQGPPERDLC

       190        200        210        220        230        240
         |          |          |          |          |          |
GCYSVSSVLP GCAEPWNHGK TFTCTAAYPE SKTPLTATLS KSGNTFRPEV HLLPPPSEEL

       250        260        270        280        290        300
         |          |          |          |          |          |
ALNELVTLTC LARGFSPKDV LVRWLQGSQE LPREKYLTWA SRQEPSQGTT TFAVTSILRV

       310        320        330        340        350
         |          |          |          |          |
AAEDWKKGDT FSCMVGHEAL PLAFTQKTID RLAGKPTHVN VSVVMAEVDG TCY
```

CODING SEQUENCE

```
                                                                g        -1
     catccccgac cagccccaag gtcttcccgc tgagcctctg cagcacccag ccagatggga      60
     acgtggtcat cgcctgcctg gtccagggct tcttccccca ggagccactc agtgtgacct     120
     ggagcgaaag cggacagggc gtgaccgcca gaaacttccc acccagccag gatgcctccg     180
     gggacctgta caccacgagc agccagctga ccctgccggc cacacagtgc ctagccggca     240
     agtccgtgac atgccacgtg aagcactaca cgaatcccag ccaggatgtg actgtgccct     300
     gcccagttcc ctcaactcca cctaccccat ctccctcaac tccacctacc ccatctccct     360
     catgctgcca cccccgactg tcactgcacc gaccggccct cgaggacctg ctcttaggtt     420
     cagaagcgaa cctcacgtgc acactgaccg gcctgagaga tgcctcaggt gtcaccttca     480
     cctggacgcc ctcaagtggg aagagcgctg ttcaaggacc acctgagcgt gacctctgtg     540
     gctgctacag cgtgtccagt gtcctgccgg gctgtgccga gccatggaac catgggaaga     600
     ccttcacttg cactgctgcc taccccgagt ccaagacccc gctaaccgcc accctctcaa     660
     aatccggaaa cacattccgg cccgaggtcc acctgctgcc gccgccgtcg gaggagctgg     720
     ccctgaacga gctggtgacg ctgacgtgcc tggcacgcgg cttcagcccc aaggacgtgc     780
     tggttcgctg gctgcagggg tcacaggagc tgccccgcga gaagtacctg acttgggcat     840
     cccggcagga gcccagccag ggcaccacca ccttcgctgt gaccagcata ctgcgcgtgg     900
     cagccgagga ctggaagaag ggggacacct tctcctgcat ggtgggccac gaggccctgc     960
     cgctggcctt cacacagaag accatcgacc gcttggcggg taaacccacc catgtcaatg    1020
     tgtctgttgt catggcggag gtggacggca cctgctactg a                       1061
```

GenBank

```
J00220
LOCUS       HUMIGCC8     2533 bp    DNA             PRI       02-DEC-1998
DEFINITION  Homo sapiens immunoglobulin alpha-1 heavy chain constant region
            (IGHA1) gene, partial cds.
ACCESSION   J00220
VERSION     J00220.1  GI:184743
KEYWORDS    .
SOURCE      human.
```

FIGURE 7 B

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2533)
  AUTHORS   Takahashi,N., Ueda,S., Obata,M., Nikaido,T., Nakai,S. and Honjo,T.
  TITLE     Structure of human immunoglobulin gamma genes: Implications for
            evolution of a gene family
  JOURNAL   Cell 29, 671-679 (1982)
  MEDLINE   83001943
COMMENT     This sequence is part of a multigene region containing the
            immunoglobulin heavy chain gamma-3, gamma-1, pseudo-epsilon, and
            alpha-1 genes.
FEATURES             Location/Qualifiers
     source          1..2533
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q32.33"
                     /clone="cosmid Ig13; Ig-gamma3-122"
                     /tissue_type="placenta; liver"
                     /germline
     intron          <1..141
                     /note="alpha-1 intron J-C"
     exon            142..447
                     /gene="IGHA1"
     gene            <142..>1638
                     /gene="IGHA1"
     CDS             join(<142..447,662..1021,1244..1638)
                     /gene="IGHA1"
                     /codon_start=3
                     /product="immunoglobulin alpha-1 heavy chain constant
                     region"
                     /protein_id="AAC82528.1"
                     /db_xref="GI:184749"
                     /translation="SPTSPKVFPLSLCSTQPDGNVVIACLVQGFFPQEPLSVTWSESG
                     QGVTARNFPPSQDASGDLYTTSSQLTLPATQCLAGKSVTCHVKHYTNPSQDVTVPCPV
                     PSTPPTPSPSTPPTPSPSCCHPRLSLHRPALEDLLLGSEANLTCTLTGLRDASGVTFT
                     WTPSSGKSAVQGPPERDLCGCYSVSSVLPGCAEPWNHGKTFTCTAAYPESKTPLTATL
                     SKSGNTFRPEVHLLPPPSEELALNELVTLTCLARGFSPKDVLVRWLQGSQELPREKYL
                     TWASRQEPSQGTTTFAVTSILRVAAEDWKKGDTFSCMVGHEALPLAFTQKTIDRLAGK
                     PTHVNVSVVMAEVDGTCY"
     intron          448..661
                     /gene="IGHA1"
                     /note="A"
     exon            662..1021
                     /gene="IGHA1"
     intron          1022..1243
                     /gene="IGHA1"
                     /note="B"
     exon            1244..>1638
                     /gene="IGHA1"
BASE COUNT      490 a    866 c    753 g    424 t
ORIGIN
        1 ggtccaactg caggcctgtg gtgcaggagc tgtgtgacca tggggctgtc accaggcctc
       61 tctgtgctgg gttcctccag tatagaggag aggcagtata gaggagaggg ccgcgtcctc
      121 acagtgcatt ctgtgttcca gcatccccga ccagccccaa ggtcttcccg ctgagcctct
      181 gcagcaccca gccagatggg aacgtggtca tcgcctgcct ggtccagggc ttcttccccc
      241 aggagccact cagtgtgacc tggagcgaaa gcggacaggg cgtgaccgcc agaaacttcc
      301 cacccagcca ggatgcctcc ggggacctgt acaccacgag cagccagctg accctgccgg
      361 ccacacagtg cctagccggc aagtccgtga catgccacgt gaagcactac acgaatccca
      421 gccaggatgt gactgtgccc tgcccaggtc agagggcagg ctggggagtg gggcggggcc
      481 accccgtcgt gccctgacac tgcgcctgca cccgtgttcc ccacagggag ccgccccttc
      541 actcacacca gagtggaccc cgggccgagc cccaggaggt ggtggtggac aggccaggag
```

FIGURE 7 C

```
 601 gggcgaggcg ggggcatggg gaagtatgtg ctgaccagct caggccatct ctccactcca
 661 gttccctcaa ctccacctac cccatctccc tcaactccac ctaccccatc tccctcatgc
 721 tgccaccccc gactgtcact gcaccgaccg gccctcgagg acctgctctt aggttcagaa
 781 gcgaacctca cgtgcacact gaccggcctg agagatgcct caggtgtcac cttcacctgg
 841 acgccctcaa gtgggaagag cgctgttcaa ggaccacctg agcgtgacct ctgtggctgc
 901 tacagcgtgt ccagtgtcct gccgggctgt gccgagccat ggaaccatgg gaagaccttc
 961 acttgcactg ctgcctaccc cgagtccaag accccgctaa ccgccaccct ctcaaaatcc
1021 ggtgggtcca gaccctgctc ggggccctgc tcagtgctct ggtttgcaaa gcatattcct
1081 ggcctgcctc ctccctccca atcctgggct ccagtgctca tgccaagtac acagggaaac
1141 tgaggcaggc tgaggggcca ggacacagcc cggggtgccc accagagcag aggggctctc
1201 tcatcccctg cccagccccc tgacctggct ctctaccctc caggaaacac attccggccc
1261 gaggtccacc tgctgccgcc gccgtcggag gagctggccc tgaacgagct ggtgacgctg
1321 acgtgcctgg cacgcggctt cagccccaag gacgtgctgg ttcgctggct gcaggggtca
1381 caggagctgc cccgcgagaa gtacctgact tgggcatccc ggcaggagcc cagccagggc
1441 accaccacct tcgctgtgac cagcatactg cgcgtggcag ccgaggactg gaagaagggg
1501 gacaccttct cctgcatggt gggccacgag gccctgccgc tggccttcac acagaagacc
1561 atcgaccgct tggcgggtaa acccacccat gtcaatgtgt ctgttgtcat ggcggaggtg
1621 gacggcacct gctactgagc cgcccgcctg tccccacccc tgaataaact ccatgctccc
1681 ccaagcagcc ccacgcttcc atccggcgcc tgtctgtcca tcctcagggt ctcagcactt
1741 gggaaagggc cagggcatgg acagggaaga ataccccctg ccctgagcct cgggggccc
1801 ctggcacccc catgagactt tccaccctgg tgtgagtgtg agttgtgagt gtgagagtgt
1861 gtggtgcagg aggcctcgct ggtgtgagat cttaggtctg ccaaggcagg cacagcccag
1921 gatgggttct gagagacgca catgccccgg acagttctga gtgagcagtg gcatggccgt
1981 ttgtccctga gagagccgcc tctggctgta gctgggaggg aataggggag gtaaaaggag
2041 caggctagcc aagaaaggcg caggtagtgg caggagcggc gagggagtga ggggctggac
2101 tccagggccc cactgggagg acaagctcca ggagggcccc accaccctag tgggtgggcc
2161 tcaggacgtc ccactgacgc atgcaggaag gggcacctcc ccttaaccac actgctctgt
2221 acggggcacg tgggcacagg tgcacactca cactcacata tatgcctgag ccctgcagga
2281 gcggaacgtt cacagcccag acccagttcc agaaaagcca ggggagtccc ctcccaagcc
2341 cccaagctca gcctgctccc ctaggccccc ctggcttccc tgtgtttcca ctgtgcacag
2401 atcaggcacc aactccacag acccctccca ggcagccccc gctccctgcc tggccaagtc
2461 tcccatcccc tcctaagccc aactaggacc caaagcatag acagggaggg gccacgtggg
2521 gtggcatcag aag
```

II. HUMAN IG ALPHA-2 CHAIN C REGION - HOMO SAPIENS (HUMAN).

AMINO ACID SEQUENCE

>sp|P01877|ALC2_HUMAN IG ALPHA-2 CHAIN C REGION - Homo sapiens (Human).

```
        10         20         30         40         50         60
         |          |          |          |          |          |
ASPTSPKVFP LSLDSTPQDG NVVVACLVQG FFPQEPLSVT WSESGQNVTA RNFPPSQDAS

        70         80         90        100        110        120
         |          |          |          |          |          |
GDLYTTSSQL TLPATQCPDG KSVTCHVKHY TNPSQDVTVP CPVPPPPPCC HPRLSLHRPA

       130        140        150        160        170        180
         |          |          |          |          |          |
LEDLLLGSEA NLTCTLTGLR DASGATFTWT PSSGKSAVQG PPERDLCGCY SVSSVLPGCA

       190        200        210        220        230        240
         |          |          |          |          |          |
QPWNHGETFT CTAAHPELKT PLTANITKSG NTFRPEVHLL PPPSEELALN ELVTLTCLAR

       250        260        270        280        290        300
         |          |          |          |          |          |
GFSPKDVLVR WLQGSQELPR EKYLTWASRQ EPSQGTTTFA VTSILRVAAE DWKKGDTFSC
```

FIGURE 7 D

```
        310         320         330         340
         |           |           |           |
MVGHEALPLA FTQKTIDRLA GKPTHVNVSV VMAEVDGTCY
```

CODING SEQUENCE

```
                                                                      g       -1
catccccgac cagccccaag gtcttcccgc tgagcctcga cagcaccccc caagatggga      60
acgtggtcgt cgcatgcctg gtccagggct tcttccccca ggagccactc agtgtgacct     120
ggagcgaaag cggacagaac gtgaccgcca gaaacttccc acctagccag gatgcctccg     180
gggacctgta caccacgagc agccagctga ccctgccggc cacacagtgc ccagacggca     240
agtccgtgac atgccacgtg aagcactaca cgaatcccag ccaggatgtg actgtgccct     300
gcccagttcc cccacctccc ccatgctgcc acccccgact gtcgctgcac cgaccggccc     360
tcgaggacct gctcttaggt tcagaagcga acctcacgtg cacactgacc ggcctgagag     420
atgcctctgg tgccaccttc acctggacgc cctcaagtgg gaagagcgct gttcaaggac     480
cacctgagcg tgacctctgt ggctgctaca gcgtgtccag tgtcctgcct ggctgtgccc     540
agccatggaa ccatggggag accttcacct gcactgctgc ccaccccgag ttgaagaccc     600
cactaaccgc caacatcaca aaatccggaa acacattccg gcccgaggtc cacctgctgc     660
cgccgccgtc ggaggagctg gccctgaacg agctggtgac gctgacgtgc ctggcacgtg     720
gcttcagccc caaggatgtg ctggttcgct ggctgcaggg gtcacaggag ctgccccgcg     780
agaagtacct gacttgggca tcccggcagg agcccagcca gggcaccacc accttcgctg     840
tgaccagcat actgcgcgtg gcagccgagg actggaagaa gggggacacc ttcтcctgca     900
tggtgggcca cgaggccctg ccgctggcct tcacacagaa gaccatcgac cgcttggcgg     960
gtaaacccac ccatgtcaat gtgtctgttg tcatggcgga ggtggacggc acctgctact    1020
ga                                                                   1022
```

GenBank

```
J00221      Human Ig germline
LOCUS       HUMIGCD7     2516 bp    DNA             PRI       11-APR-2001
DEFINITION  Human Ig germline H-chain G-E-A region B: alpha-2 A2m(1) allele
            constant region, 3' end.
ACCESSION   J00221
VERSION     J00221.1  GI:184756
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2516)
  AUTHORS   Ellison,J. and Hood,L.
  TITLE     Linkage and sequence homology of two human immunoglobulin gamma
            heavy chain constant region genes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 79 (6), 1984-1988 (1982)
  MEDLINE   82197621
   PUBMED   6804948
REFERENCE   2  (bases 737 to 1016)
  AUTHORS   Flanagan,J.G. and Rabbitts,T.H.
  TITLE     Arrangement of human immunoglobulin heavy chain constant region
            genes implies evolutionary duplication of a segment containing
            gamma, epsilon and alpha genes
  JOURNAL   Nature 300 (5894), 709-713 (1982)
  MEDLINE   83088998
   PUBMED   6817141
REFERENCE   3  (bases 49 to 229; 425 to 514)
  AUTHORS   Hisajima,H., Nishida,Y., Nakai,S., Takahashi,N., Ueda,S. and
            Honjo,T.
  TITLE     Structure of the human immunoglobulin C epsilon 2 gene, a truncated
            pseudogene: implications for its evolutionary origin
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 80 (10), 2995-2999 (1983)
  MEDLINE   83221529
   PUBMED   6407005
REFERENCE   4  (bases 1 to 2516)
  AUTHORS   Flanagan,J.G., Lefranc,M.P. and Rabbitts,T.H.
```

FIGURE 7 E

```
  TITLE     Mechanisms of divergence and convergence of the human
            immunoglobulin alpha 1 and alpha 2 constant region gene sequences
  JOURNAL   Cell 36 (3), 681-688 (1984)
  MEDLINE   84130179
   PUBMED   6421489
COMMENT     [3] also reports the complete alpha-1 gene and part of the A2m(2)
            alpha-2 allele (bases 737-2516; see Sites table). Comparison of the
            three sequences suggests that the A2m(1) alpha-2 allele might be a
            hybrid of the alpha-1 gene and A2m(2) alpha-2 allele. The hinge
            region in the alpha genes occurs at beginning of the CH2 domain.
            The alpha-1 hinge region is 13 amino acids longer than that in
            alpha-2. Both hinge regions consist of approximate tandem repeats
            of a 15 bp sequence. The first repeat occurs 5' to the mRNA splice
            site and is non-coding. The authors [3] suggest that this
            repetitive structure provides a possible mechanism for the large
            number of variations observed in hinge regions. There is a coupled
            30 bp insertion, 9 bp deletion in alpha-2 relative to alpha-1
            (starting at base 97).
            [1] also reports sequences for the epsilon-1 and epsilon-2
            (pseudogene) C-region genes. The authors [1] determined the
            physical linkage between epsilon-1 and alpha-2 and that between
            epsilon-2 and alpha-1. [2] also reports the alpha-1 CH2 domain and
            epsilon-2.
            This entry is part of a multigene region (region B), which includes
            the gamma-2, gamma-4, epsilon-1 and alpha-2 genes. See segment 1
            for more comments.
            Complete source information:
            Human genomic DNA, cosmid Ig10 [1],[3]; placenta DNA [2] clone
            H-Ig-alpha-25; genomic DNA from TOU II-5 library clone
            lambda-TOU-alpha2 (for A2m(2) allele) [3].
FEATURES              Location/Qualifiers
     source           1..2516
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /map="14q32.33"
                      /germline
     gene             <1..1621
                      /gene="IgH"
                      /note="IGHA2"
     intron           <1..163
                      /gene="IgH"
                      /note="alpha-2 intron J-C"
     CDS              join(<164..469,684..1004,1227..1621)
                      /gene="IgH"
                      /note="contains constant region"
                      /codon_start=3
                      /product="immunoglobulin alpha-2 heavy chain"
                      /protein_id="AAB59396.1"
                      /db_xref="GI:184761"
                      /translation="SPTSPKVFPLSLDSTPQDGNVVVACLVQGFFPQEPLSVTWSESG
                      QNVTARNFPPSQDASGDLYTTSSQLTLPATQCPDGKSVTCHVKHYTNPSQDVTVPCPV
                      PPPPPCCHPRLSLHRPALEDLLLGSEANLTCTLTGLRDASGATFTWTPSSGKSAVQGP
                      PERDLCGCYSVSSVLPGCAQPWNHGETFTCTAAHPELKTPLTANITKSGNTFRPEVHL
                      LPPPSEELALNELVTLTCLARGFSPKDVLVRWLQGSQELPREKYLTWASRQEPSQGTT
                      TFAVTSILRVAAEDWKKGDTFSCMVGHEALPLAFTQKTIDRLAGKPTHVNVSVVMAEV
                      DGTCY"
     exon             164..469
                      /gene="IgH"
                      /note="G00-119-333"
     intron           470..683
                      /gene="IgH"
                      /note="alpha-2 intron A"
     exon             684..1004
```

FIGURE 7 F

```
                     /gene="IgH"
     intron          1005..1226
                     /gene="IgH"
                     /note="alpha-2 intron B"
     exon            1227..1621
                     /gene="IgH"
     variation       1434
                     /gene="IgH"
                     /note="t in A2m(1); a in A2m(2)"
     variation       1441
                     /gene="IgH"
                     /note="g in A2m(1); a in A2m(2)"
     variation       1465
                     /gene="IgH"
                     /note="c in A2m(1); t in A2m(2)"
     variation       1486
                     /gene="IgH"
                     /note="c in A2m(1); g in A2m(2)"
     variation       1553
                     /gene="IgH"
                     /note="t in A2m(1); a in A2m(2)"
     variation       1573..1574
                     /gene="IgH"
                     /note="tg in A2m(1); ca in A2m(2)"
     variation       1602..1606
                     /gene="IgH"
                     /note="tggac in A2m(1); cggat in A2m(2)"
     variation       2060
                     /note="c in A2m(1); t in A2m(2)"
     variation       2384
                     /note="a in A2m(1); c in A2m(2)"
     variation       2390
                     /note="c in A2m(1); g in A2m(2)"
BASE COUNT      488 a    861 c    754 g    413 t
ORIGIN
        1 ggtccaaccg caggcccatg gtgcaggagc tgtgtaacct atggggctgt caccaggcct
       61 ctctgtgctg ggttcctcca gtgtagagga gaggcaggta cagcctgtcc tcctggggac
      121 atggcatgag ggccgcgtcc tcacagcgca ttctgtgttc cagcatcccc gaccagcccc
      181 aaggtcttcc cgctgagcct cgacagcacc ccccaagatg ggaacgtggt cgtcgcatgc
      241 ctggtccagg gcttcttccc ccaggagcca ctcagtgtga cctggagcga aagcggacag
      301 aacgtgaccg ccagaaactt cccacctagc caggatgcct ccggggacct gtacaccacg
      361 agcagccagc tgaccctgcc ggccacacag tgcccagacg gcaagtccgt gacatgccac
      421 gtgaagcact acacgaatcc cagccaggat gtgactgtgc cctgcccagg tcagagggca
      481 ggctggggag tggggcgggg ccaccccgtc ctgccctgac actgcgcctg cacccgtgtt
      541 ccccacaggg agccgccccc tcactcacac cagagtggac cccgggccga gccccaggag
      601 gtggtggtgg acaggccagg aggggcgagg cgggggcacg gggaagggcg ttctgaccag
      661 ctcaggccat ctctccactc cagttccccc acctcccccа tgctgccacc cccgactgtc
      721 gctgcaccga ccggccctcg aggacctgct cttaggttca gaagcgaacc tcacgtgcac
      781 actgaccggc ctgagagatg cctctggtgc caccttcacc tggacgccct caagtgggaa
      841 gagcgctgtt caaggaccac ctgagcgtga cctctgtggc tgctacagcg tgtccagtgt
      901 cctgcctggc tgtgcccagc catggaacca tggggagacc ttcacctgca ctgctgccca
      961 ccccgagttg aagaccccac taaccgccaa catcacaaaa tccggtgggt ccagaccctg
     1021 ctcggggccc tgctcagtgc tctggtttgc aaagcatatt cccggcctgc ctcctccctc
     1081 ccaatcctgg gctccagtgc tcatgccaag tacacaggga aactgaggca ggctgagggg
     1141 ccaggacaca gcccagggtg cccaccagag cagaggggct ctctcatccc ctgcccagcc
     1201 ccctgacctg gctctctacc ctccaggaaa cacattccgg cccgaggtcc acctgctgcc
     1261 gccgccgtcg gaggagctgg ccctgaacga gctggtgacg ctgacgtgcc tggcacgtgg
     1321 cttcagcccc aaggatgtgc tggttcgctg gctgcagggg tcacaggagc tgccccgcga
     1381 gaagtacctg actttgggcat cccggcagga gcccagccag ggcaccacca ccttcgctgt
     1441 gaccagcata ctgcgcgtgg cagccgagga ctggaagaag ggggacacct tctcctgcat
     1501 ggtgggccac gaggccctgc cgctggcctt cacacagaag accatcgacc gcttggcggg
     1561 taaacccacc catgtcaatg tgtctgttgt catggcggag gtggacggca cctgctactg
     1621 agccgcccgc ctgtccccac ccctgaataa actccatgct cccccaagca gccccacgct
```

FIGURE 7 G

```
1681 tccatccggc gcctgtctgt ccatcctcag ggtctcagca cttgggaaag ggccagggca
1741 tggacaggga agaatacccc ctgccctgag cctcgggggg cccctggcac ccccatgaga
1801 ctttccaccc tggtgtgagt gtgagttgtg agtgtgagag tgtgtggtgc aggaggcctc
1861 gctggtgtga gatcttaggt ctgccaaggc aggcacagcc caggatgggt tctgagagac
1921 gcacatgccc cggacagttc tgagtgagca gtggcatggc cgtttgtccc tgagagagcc
1981 gcctctggct gtagctggga gggaataggg agggtaaaag gagcaggcta gccaagaaag
2041 gcgcaggtag tggcaggagc ggcgagggag tgaggggctg gactccaggg ccccactggg
2101 aggacaagct ccaggagggc cccaccaccc tagtgggtgg gcctcaggac gtcccactga
2161 cgcatgcagg aaggggcacc tccccttaac cacactgctc tgtacggggc acgtgggcac
2221 acatgcacac tcacactcac atatacgcct gagccctgca ggagtggaac gttcacagcc
2281 cagacccagt tccagaaaag ccaggggagt cccctcccaa gccccaagc tcagcctgct
2341 cccccaggcc cctctggctt ccctgtgttt ccactgtgca cagatcaggc accaactcca
2401 cagacccctc ccaggcagcc cctgctccct gcctggccaa gtctcccatc ccttcctaag
2461 cccaactagg acccaaagca tagacaggga ggggccgcgt ggggtggcat cagaag
```

III. HUMAN IG GAMMA-1 CHAIN C REGION - HOMO SAPIENS (HUMAN)

AMINO ACID SEQUENCE

>sp|P01857|GC1_HUMAN IG GAMMA-1 CHAIN C REGION - Homo sapiens (Human).

```
         10         20         30         40         50         60
          |          |          |          |          |          |
ASTKGPSVFP LAPSSKSTSG GTAALGCLVK DYFPEPVTVS WNSGALTSGV HTFPAVLQSS

         70         80         90        100        110        120
          |          |          |          |          |          |
GLYSLSSVVT VPSSSLGTQT YICNVNHKPS NTKVDKKVEP KSCDKTHTCP PCPAPELLGG

        130        140        150        160        170        180
          |          |          |          |          |          |
PSVFLFPPKP KDTLMISRTP EVTCVVVDVS HEDPEVKFNW YVDGVEVHNA KTKPREEQYN

        190        200        210        220        230        240
          |          |          |          |          |          |
STYRVVSVLT VLHQDWLNGK EYKCKVSNKA LPAPIEKTIS KAKGQPREPQ VYTLPPSRDE

        250        260        270        280        290        300
          |          |          |          |          |          |
LTKNQVSLTC LVKGFYPSDI AVEWESNGQP ENNYKITPPV LDSDGSFFLY SKLTVDKSRW

        310        320        330
          |          |          |
QQGNVFSCSV MHEALHNHYT QKSLSLSPGK
```

CODING SEQUENCE

```
                                                                  g        -1
cctccaccaa gggcccatcg gtcttccccc tggcaccctc ctccaagagc acctctgggg         60
gcacagcggc cctgggctgc ctggtcaagg actacttccc cgaaccggtg acggtgtcgt        120
ggaactcagg cgccctgacc agcggcgtgc acaccttccc ggctgtccta cagtcctcag        180
gactctactc cctcagcagc gtggtgaccg tgccctccag cagcttgggc acccagacct        240
acatctgcaa cgtgaatcac aagcccagca acaccaaggt ggacaagaaa gttgagccca        300
aatcttgtga caaaactcac acatgcccac cgtgcccagc acctgaactc ctggggggac        360
cgtcagtctt cctcttcccc ccaaaaccca aggacaccct catgatctcc cggacccctg        420
aggtcacatg cgtggtggtg gacgtgagcc acgaagaccc tgaggtcaag ttcaactggt        480
acgtggacgg cgtggaggtg cataatgcca agacaaagcc gcgggaggag cagtacaaca        540
gcacgtaccg ggtggtcagc gtcctcaccg tcctgcacca ggactggctg aatggcaagg        600
agtacaagtg caaggtctcc aacaaagccc tcccagcccc catcgagaaa accatctcca        660
aagccaaagg gcagccccga gaaccacagg tgtacaccct gccccatcc cgggatgagc        720
```

FIGURE 7 H

```
tgaccaagaa ccaggtcagc ctgacctgcc tggtcaaagg cttctatccc agcgacatcg      780
ccgtggagtg ggagagcaat gggcagccgg agaacaacta caagaccacg cctcccgtgc      840
tggactccga cggctccttc ttcctctaca gcaagctcac cgtggacaag agcaggtggc      900
agcaggggaa cgtcttctca tgctccgtga tgcatgaggc tctgcacaac cactacacgc      960
agaagagcct ctccctgtct ccgggtaaat ga                                   992
```

GenBank

```
J00228.
LOCUS       HUMIGCC4     2009 bp    DNA             PRI       02-DEC-1998
DEFINITION  Homo sapiens immunoglobulin gamma-1 heavy chain constant region
            (IGHG1) gene, partial cds.
ACCESSION   J00228
VERSION     J00228.1  GI:184739
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2009)
  AUTHORS   Takahashi,N., Ueda,S., Obata,M., Nikaido,T., Nakai,S. and Honjo,T.
  TITLE     Structure of human immunoglobulin gamma genes: Implications for
            evolution of a gene family
  JOURNAL   Cell 29, 671-679 (1982)
  MEDLINE   83001943
COMMENT     This sequence is part of a multigene region containing the
            immunoglobulin heavy chain gamma-3, gamma-1, pseudo-epsilon, and
            alpha-1 genes.
FEATURES             Location/Qualifiers
     source          1..2009
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q32.33"
                     /clone="cosmid Ig13; Ig-gamma3-122"
                     /tissue_type="placenta; liver"
                     /germline
     gene            <1..>1803
                     /gene="IGHG1"
     intron          <1..209
                     /gene="IGHG1"
     CDS             join(<210..503,892..936,1055..1384,1481..1803)
                     /gene="IGHG1"
                     /codon_start=3
                     /product="immunoglobulin gamma-1 heavy chain constant
                     region"
                     /protein_id="AAC82527.1"
                     /db_xref="GI:184747"
                     /translation="STKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGA
                     LTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSC
                     DKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWY
                     VDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTI
                     SKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTT
                     PPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK"
     misc_difference 563
                     /gene="IGHG1"
                     /replace=""
     misc_difference 593
                     /gene="IGHG1"
                     /replace=""
     misc_difference 614
                     /gene="IGHG1"
                     /replace=""
```

FIGURE 7

```
     misc_difference 633
                     /gene="IGHG1"
                     /replace=""
     misc_difference 643
                     /gene="IGHG1"
                     /replace=""
     misc_difference 654
                     /gene="IGHG1"
                     /replace=""
     misc_difference 684
                     /gene="IGHG1"
                     /replace=""
     misc_difference 692
                     /gene="IGHG1"
                     /replace=""
     misc_difference 765..766
                     /gene="IGHG1"
                     /replace=""
     misc_difference 1475
                     /gene="IGHG1"
                     /replace=""
     misc_difference 1578
                     /gene="IGHG1"
                     /replace=""
BASE COUNT      418 a    698 c    566 g    327 t
ORIGIN
        1 agctttctgg ggcaggccag gcctgacctt ggctttgggg cagggagggg gctaaggtga
       61 ggcaggtggc gccagcaggt gcacacccaa tgcccatgag cccagacact ggacgctgaa
      121 cctcgcggac agttaagaac ccaggggcct ctgcgcctgg gcccagctct gtcccacacc
      181 gcggtcacat ggcaccacct ctcttgcagc ctccaccaag ggcccatcgg tcttcccct
      241 ggcaccctcc tccaagagca cctctggggg cacagcggcc ctgggctgcc tggtcaagga
      301 ctacttcccc gaaccggtga cggtgtcgtg gaactcaggc gccctgacca gcggcgtgca
      361 caccttcccg gctgtcctac agtcctcagg actctactcc ctcagcagcg tggtgaccgt
      421 gccctccagc agcttgggca cccagaccta catctgcaac gtgaatcaca agcccagcaa
      481 caccaaggtg gacaagaaag ttggtgagag gccagcacag ggagggaggg tgtctgctgg
      541 aagcaggctc agcgctcctg cctggacgca tcccggctat gcagccccag tccagggcag
      601 caaggcaggc cccgtctgcc tcttcacccg gagcctctgc ccgccccact catgctcagg
      661 gagagggtct tctggctttt tcccaggctc tgggcaggca caggctaggt gccccctaacc
      721 caggccctgc acacaaaggg gcaggtgctg ggctcagacc tgccaagagc catatccggg
      781 aggaccctgc ccctgaccta agcccacccc aaaggccaaa ctctccactc cctcagctcg
      841 gacaccttct ctcctcccag attccagtaa ctcccaatct tctctctgca gagcccaaat
      901 cttgtgacaa aactcacaca tgcccaccgt gcccaggtaa gccagcccag gcctcgccct
      961 ccagctcaag gcgggacagg tgccctagag tagcctgcat ccagggacag gccccagccg
     1021 ggtgctgaca cgtccacctc catctcttcc tcagcacctg aactcctggg gggaccgtca
     1081 gtcttcctct tccccccaaa acccaaggac accctcatga tctcccggac ccctgaggtc
     1141 acatgcgtgg tggtggacgt gagccacgaa gaccctgagg tcaagttcaa ctggtacgtg
     1201 gacggcgtgg aggtgcataa tgccaagaca aagccgcggg aggagcagta caacagcacg
     1261 taccgggtgg tcagcgtcct caccgtcctg caccaggact ggctgaatgg caaggagtac
     1321 aagtgcaagg tctccaacaa agccctccca gccccatcg agaaaaccat ctccaaagcc
     1381 aaaggtggga cccgtggggt gcgagggcca catggacaga ggccggctcg gcccaccctc
     1441 tgccctgaga gtgaccgctg taccaacctc tgtcctacag ggcagccccg agaaccacag
     1501 gtgtacaccc tgcccccatc ccgggatgag ctgaccaaga accaggtcag cctgacctgc
     1561 ctggtcaaag gcttctatcc cagcgacatc gccgtggagt gggagagcaa tgggcagccg
     1621 gagaacaact acaagaccac gcctcccgtg ctggactccg acggctcctt cttcctctac
     1681 agcaagctca ccgtggacaa gagcaggtgg cagcagggga acgtcttctc atgctccgtg
     1741 atgcatgagg ctctgcacaa ccactacacg cagaagagcc tctccctgtc tccgggtaaa
     1801 tgagtgcgac ggccggcaag ccccgctccc cgggctctcg cggtcgcacg aggatgcttg
     1861 gcacgtaccc cctgtacata cttcccgggc gcccagcatg gaaataaagc acccagcgct
     1921 gccctgggcc cctgcgagac tgtgatggtt ctttccacgg gtcaggccga gtctgaggcc
     1981 tgagtggcat gagggaggca gagcgggtc
```

FIGURE 7 J

IV. HUMAN IG GAMMA-2 CHAIN C REGION - HOMO SAPIENS (HUMAN).

AMINO ACID SEQUENCE

>sp|P01859|GC2_HUMAN IG GAMMA-2 CHAIN C REGION - Homo sapiens (Human).

```
        10         20         30         40         50         60
         |          |          |          |          |          |
ASTKGPSVFP LAPCSRSTSE STAALGCLVK DYFPEPVTVS WNSGALTSGV HTFPAVLQSS

        70         80         90        100        110        120
         |          |          |          |          |          |
GLYSLSSVVT VPSSNFGTQT YTCNVDHKPS NTKVDKTVER KCCVECPPCP APPVAGPSVF

       130        140        150        160        170        180
         |          |          |          |          |          |
LFPPKPKDTL MISRTPEVTC VVVDVSHEDP EVQFNWYVDG VEVHNAKTKP REEQFNSTFR

       190        200        210        220        230        240
         |          |          |          |          |          |
VVSVLTVVHQ DWLNGKEYKC KVSNKGLPAP IEKTISKTKG QPREPQVYTL PPSREEMTKN

       250        260        270        280        290        300
         |          |          |          |          |          |
QVSLTCLVKG FYPSDIAVEW ESNGQPENNY KTTPPMLDSD GSFFLYSKLT VDKSRWQQGN

       310        320    326
         |          |      |
VFSCSVMHEA LHNHYTQKSL SLSPGK
```

CODING SEQUENCE

```
     gcctccacca agggcccatc ggtcttcccc ctggcgccct gctccaggag cacctccgag        60
     agcacagccg ccctgggctg cctggtcaag gactacttcc ccgaaccggt gacggtgtcg       120
     tggaactcag gcgctctgac cagcggcgtg cacaccttcc cagctgtcct acagtcctca       180
     ggactctact ccctcagcag cgtggtgacc gtgccctcca gcaacttcgg cacccagacc       240
     tacacctgca acgtagatca caagcccagc aacaccaagg tggacaagac agttgagcgc       300
     aaatgttgtg tcgagtgccc accgtgccca gcaccacctg tggcaggacc gtcagtcttc       360
     ctcttccccc caaaacccaa ggacaccctc atgatctccc ggacccctga ggtcacgtgc       420
     gtggtggtgg acgtgagcca cgaagaccct gaggtccagt tcaactggta cgtggacggc       480
     gtggaggtgc ataatgccaa gacaaagcca cgggaggagc agttcaacag cacgttccgt       540
     gtggtcagcg tcctcaccgt tgtgcaccag gactggctga acggcaagga gtacaagtgc       600
     aaggtctcca acaaaggcct cccagccccc atcgagaaaa ccatctccaa aaccaaaggg       660
     cagccccgag aaccacaggt gtacaccctg cccccatccc gggaggagat gaccaagaac       720
     caggtcagcc tgacctgcct ggtcaaaggc ttctacccca gcgacatcgc cgtggagtgg       780
     gagagcaatg ggcagccgga gaacaactac aagaccacac ctcccatgct ggactccgac       840
     ggctccttct tcctctacag caagctcacc gtggacaaga gcaggtggca gcaggggaac       900
     gtcttctcat gctccgtgat gcatgaggct ctgcacaacc actacacgca gaagagcctc       960
     tccctgtctc cgggtaaa                                                    978
```

GenBank

```
J00230. Human Ig germline ...
LOCUS       HUMIGCD1     2009 bp    DNA             PRI       11-APR-2001
DEFINITION  Human Ig germline H-chain G-E-A region B: gamma-2 constant region,
            3' end.
ACCESSION   J00230 V00554
VERSION     J00230.1  GI:184750
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

FIGURE 7 K

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2009)
  AUTHORS   Ellison,J. and Hood,L.
  TITLE     Linkage and sequence homology of two human immunoglobulin gamma
            heavy chain constant region genes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 79 (6), 1984-1988 (1982)
  MEDLINE   82197621
   PUBMED   6804948
REFERENCE   2  (bases 896 to 1256; 1749 to 1937)
  AUTHORS   Krawinkel,U. and Rabbitts,T.H.
  TITLE     Comparison of the hinge-coding segments in human immunoglobulin
            gamma heavy chain genes and the linkage of the gamma 2 and gamma 4
            subclass genes  JOURNAL   EMBO J. 1 (4), 403-407 (1982)
  MEDLINE   84235992   PUBMED   6329676
REFERENCE   3  (bases 475 to 1071; 1179 to 1330; 1461 to 1524)
  AUTHORS   Takahashi,N., Ueda,S., Obata,M., Nikaido,T., Nakai,S. and Honjo,T.
  TITLE     Structure of human immunoglobulin gamma genes: implications for
            evolution of a gene family  JOURNAL   Cell 29 (2), 671-679 (1982)
  MEDLINE   83001943   PUBMED   6811139
COMMENT     On Mar 2, 2000 this sequence version replaced gi:32759.
            [2] also reports sequences for gamma-3, gamma-4, and a gamma
            pseudogene. Most of this sequence is 95% homologous with gamma-4.
            The hinge exons are only 70% homologous. The authors estimate that
            gamma-2 and gamma-4 diverged 6.6 million years ago. The authors in
            [1] speculate that intron-mediated domain transfer played an
            important role in the evolution of human gamma genes. They also
            report the hinge regions of gamma-1, gamma-3, gamma-4, and a
            pseudo-gamma gene. [1] estimates the divergence of the human gamma
            genes to be between 7.7 and 4.4 million years ago. This entry is
            part of a multigene region containing the gamma-2, gamma-4,
            epsilon-1, and alpha-2 genes. The relative locations of the four
            genes were determined by Flanagan and Rabbitts (Nature 300, 709-713
            (1982)). They refer to this gene group as region B. The region A
            genes are gamma-3, gamma-1, pseudo-epsilon, alpha-1. Flanagan and
            Rabbits also determined the general locations of the two regions.
            They place region A between the JH/mu/delta region and region B.
            Complete source information:
            Human fetal liver DNA, library of T. Maniatis [3] and Lawn et al
            [2],[1]; clones p-gamma-2RPA3 [2], 5A [3], and Ig-gamma-2-15 [1].
FEATURES            Location/Qualifiers     source          1..2009
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"                   /map="14q32.33"
                    /germline     intron          <1..215
                    /gene="IgH"     gene            <1..2009
                    /gene="IgH"                     /note="IGHG2"
     exon           216..509                    /gene="IgH"
                    /note="G00-119-338"     C_region        216..508
                    /gene="IgH"
                    /note="immunoglobulin heavy chain constant region CH1"
     CDS            join(<216..509,902..937,1056..1382,1480..1802)
                    /gene="IgH"                     /codon_start=3
                    /product="immunoglobulin gamma-2 heavy chain"
                    /protein_id="AAB59393.1"
                    /db_xref="GI:184758"
                    /translation="STKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGA
                    LTSGVHTFPAVLQSSGLYSLSSVVTVPSSNFGTQTYTCNVDHKPSNTKVDKTVERKCC
                    VECPPCPAPPVAGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVQFNWYVDGV
                    EVHNAKTKPREEQFNSTFRVVSVLTVVHQDWLNGKEYKCKVSNKGLPAPIEKTISKTK
                    GQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPML
                    DSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK"
     intron         510..901                    /gene="IgH"
     conflict       537                 /gene="IgH"
                    /citation=[3]                   /replace=""
```

FIGURE 7

```
     conflict        550..551                    /gene="IgH"
                     /citation=[3]                    /replace=""
     conflict        570                       /gene="IgH"
                     /citation=[3]                    /replace=""
     conflict        777..778                  /gene="IgH"
                     /citation=[3]                    /replace=""
     conflict        791                       /gene="IgH"
                     /citation=[3]                    /replace=""
     conflict        864                       /gene="IgH"
                     /citation=[3]                    /replace=""
     C_region        901..936                  /gene="IgH"
                     /note="immunoglobulin heavy chain hinge"
     exon            902..937                  /gene="IgH"
                     /note="G00-119-338"     intron          938..1055
                     /gene="IgH"     C_region        1055..1381
                     /gene="IgH"
                     /note="immunoglobulin heavy chain constant region CH2"
     exon            1056..1382                /gene="IgH"
                     /note="G00-119-338"     intron          1383..1479
                     /gene="IgH"     C_region        1479..1799
                     /gene="IgH"
                     /note="immunoglobulin heavy chain constant region CH3"
     exon            1480..1802                /gene="IgH"
                     /note="G00-119-338"     conflict        1493
                     /gene="IgH"                       /citation=[3]
                     /replace=""     conflict        1802..1806
                     /gene="IgH"                       /citation=[2]
                     /replace=""     conflict        1814..1815
                     /gene="IgH"                       /citation=[2]
                     /replace=""     conflict        1825
                     /gene="IgH"                       /citation=[2]
                     /replace=""     conflict        1844..1853
                     /gene="IgH"                       /citation=[2]
                     /replace=""     conflict        1890
                     /gene="IgH"                       /citation=[2]
                     /replace=""     polyA_signal    1903..1908
                     /gene="IgH"     conflict        1909..1918
                     /gene="IgH"                       /citation=[2]
                     /replace=""     conflict        1929..1932
                     /gene="IgH"                       /citation=[2]
                     /replace=""BASE COUNT      410 a    700 c    568 g    331 t
ORIGIN
        1 agctttctgg ggcgagccgg gcctgacttt ggctttgggg cagggagtgg gctaaggtga
       61 ggcaggtggc gccagccagg tgcacaccca atgcccgtga gcccagacac tggaccctgc
      121 ctggaccctc gtggatagac aagaaccgag gggcctctgc gcctgggccc agctctgtcc
      181 cacaccgcgg tcacatggca ccacctctct tgcagcctcc accaagggcc catcggtctt
      241 ccccctggcg ccctgctcca ggagcacctc cgagagcaca gccgccctgg gctgcctggt
      301 caaggactac ttccccgaac cggtgacggt gtcgtggaac tcaggcgctc tgaccagcgg
      361 cgtgcacacc ttcccagctg tcctacagtc ctcaggactc tactccctca gcagcgtggt
      421 gaccgtgccc tccagcaact tcggcaccca gacctacacc tgcaacgtag atcacaagcc
      481 cagcaacacc aaggtggaca agacagttgg tgagaggcca gctcagggag ggagggtgtc
      541 tgctggaagc caggctcagc cctcctgcct ggacgcaccc cggctgtgca gccccagccc
      601 agggcagcaa ggcaggcccc atctgtctcc tcacccggag gcctctgccc gccccactca
      661 tgctcaggga gagggtcttc tggctttttc caccaggctc caggcaggca caggctgggt
      721 gccctaccc caggccttc acacacaggg gcaggtgctt ggctcagacc tgccaaaagc
      781 catatccggg aggaccctgc ccctgaccta agccgacccc aaaggccaaa ctgtccactc
      841 cctcagctcg gacaccttct ctcctcccag atccgagtaa ctcccaatct tctctctgca
      901 gagcgcaaat gttgtgtcga gtgcccaccg tgcccaggta agccagccca ggcctcgccc
      961 tccagctcaa ggcgggacag gtgccctaga gtagcctgca tccagggaca ggccccagct
     1021 gggtgctgac acgtccacct ccatctcttc ctcagcacca cctgtggcag gaccgtcagt
     1081 cttcctcttc cccccaaaac ccaaggacac cctcatgatc tcccggaccc ctgaggtcac
     1141 gtgcgtggtg gtggacgtga gccacgaaga ccccgaggtc cagttcaact ggtacgtgga
     1201 cggcgtggag gtgcataatg ccaagacaaa gccacgggag gagcagttca acagcacgtt
```

FIGURE 7

M

```
1261 ccgtgtggtc agcgtcctca ccgttgtgca ccaggactgg ctgaacggca aggagtacaa
1321 gtgcaaggtc tccaacaaag gcctcccagc ccccatcgag aaaaccatct ccaaaaccaa
1381 aggtgggacc cgcggggtat gagggccaca tggacagagg ccggctcggc ccaccctctg
1441 ccctgggagt gaccgctgtg ccaacctctg tccctacagg gcagccccga gaaccacagg
1501 tgtacaccct gcccccatcc cgggaggaga tgaccaagaa ccaggtcagc ctgacctgcc
1561 tggtcaaagg cttctacccc agcgacatcg ccgtggagtg ggagagcaat gggcagccgg
1621 agaacaacta caagaccaca cctcccatgc tggactccga cggctccttc ttcctctaca
1681 gcaagctcac cgtggacaag agcaggtggc agcaggggaa cgtcttctca tgctccgtga
1741 tgcatgaggc tctgcacaac cactacacgc agaagagcct ctccctgtct ccgggtaaat
1801 gagtgccacg gccggcaagc ccccgctccc caggctctcg gggtcgcgtg aggatgcttg
1861 gcacgtaccc cgtgtacata cttcccaggc acccagcatg gaaataaagc acccagcgct
1921 gccctgggcc cctgcgagac tgtgatggtt ctttccgtgg gtcaggccga gtctgaggcc
1981 tgagtggcat gagggaggca gagtgggtc
```

V. HUMAN IG GAMMA-3 CHAIN C REGION - HOMO SAPIENS (HUMAN).

AMINO ACID SEQUENCE

CAA27268 C gamma 3 [Homo sapiens]

```
        10          20          30          40          50          60
         |           |           |           |           |           |
ASTKGPSVFP LAPCSRSTSG GTAALGCLVK DYFPEPVTVS WNSGALTSGV HTFPAVLQSS

        70          80          90         100         110         120
         |           |           |           |           |           |
GLYSLSSVVT VPSSSLGTQT YTCNVNHKPS NTKVDKRVEL KTPLGDTTHT CPRCPEPKSC

       130         140         150         160         170         180
         |           |           |           |           |           |
DTPPPCPRCP EPKSCDTPPP CPRCPEPKSC DTPPPCPRCP APELLGGPSV FLFPPKPKDT

       190         200         210         220         230         240
         |           |           |           |           |           |
LMISRTPEVT CVVVDVSHED PEVQFKWYVD GVEVHNAKTK PREEQYNSTF RVVSVLTVLH

       250         260         270         280         290         300
         |           |           |           |           |           |
QDWLNGKEYK CKVSNKALPA PIEKTISKTK GQPREPQVYT LPPSREEMTK NQVSLTCLVK

       310         320         330         340         350         360
         |           |           |           |           |           |
GFYPSDIAVE WESSGQPENN YNTTPPMLDS DGSFFLYSKL TVDKSRWQQG NIFSCSVMHE

       370     377
         |
ALHNRFTQKS LSLSPGK
```

CODING SEQUENCE

```
GCTTCCACCAAGGGCCCATCGGTCTTCCCCCTGGCGCCCTGCTCCAGGAGCACCTCTGGGGGCACAGCGGCCCTGGG
CTGCCTGGTCAAGGACTACTTCCCCGAACCGGTGACGGTGTCGTGGAACTCAGGCGCCCTGACCAGCGGCGTGCACA
CCTTCCCGGCTGTCCTACAGTCCTCAGGACTCTACTCCCTCAGCAGCGTGGTGACCGTGCCCTCCAGCAGCTTGGGC
ACCCAGACCTACACCTGCAACGTGAATCACAAGCCCAGCAACACCAAGGTGGACAAGAGAGTTGAGCTCAAAACCCC
ACTTGGTGACACAACTCACACATGCCCACGGTGCCCAGAGCCCAAATCTTGTGACACACCTCCCCCGTGCCCACGGT
GCCCAGAGCCCAAATCTTGTGACACACCTCCCCCATGCCCACGGTGCCCAGAGCCCAAATCTTGTGACACACCTCCC
CCGTGCCCAAGGTGCCCAGCACCTGAACTCCTGGGAGGACCGTCAGTCTTCCTCTTCCCCCCAAAACCCAAGGATAC
CCTTATGATTTCCCGGACCCCTGAGGTCACGTGCGTGGTGGTGGACGTGAGCCACGAAGACCCCGAGGTCCAGTTCA
AGTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTTCCGT
```

FIGURE 7

N

GTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGAACGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGC
CCTCCCAGCCCCCATCGAGAAAACCATCTCCAAAACCAAAGGACAGCCCCGAGAACCACAGGTGTACACCCTGCCCC
CATCCCGGGAGGAGATGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTACCCCAGCGACATCGCC
GTGGAGTGGGAGAGCAGCGGGCAGCCGGAGAACAACTACAACACCACGCCTCCCATGCTGGACTCCGACGGCTCCTT
CTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACATCTTCTCATGCTCCGTGATGCATG
AGGCTCTGCACAACCGCTTCACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGA

GenBank

X03604 Human C gamma 3 gene for IgG G3m(b) heavy chain C-region from EZZ (individual II-4 of TOU) PubMed, Protein, Related Sequences, Taxonomy, OMIM, LinkOut

```
LOCUS       HSIGGC3      2590 bp    DNA             PRI       24-NOV-1993
DEFINITION  Human C gamma 3 gene for IgG G3m(b) heavy chain C-region from EZZ
            (individual II-4 of TOU).
ACCESSION   X03604 M12958
VERSION     X03604.1  GI:33070
KEYWORDS    constant region; gamma-immunoglobulin; Ig heavy chain;
            immunoglobulin.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2590)
  AUTHORS   Huck,S., Fort,P., Crawford,D.H., Lefranc,M.P. and Lefranc,G.
  TITLE     Sequence of a human immunoglobulin gamma 3 heavy chain constant
            region gene: comparison with the other human C gamma genes
  JOURNAL   Nucleic Acids Res. 14 (4), 1779-1789 (1986)
  MEDLINE   86148507
REFERENCE   2  (bases 4 to 204; 2202 to 2236)
  AUTHORS   Takahashi,N., Ueda,S., Obata,M., Nikaido,T., Nakai,S. and
Honjo,T.
  TITLE     Structure of human immunoglobulin gamma genes: implications for
            evolution of a gene family
  JOURNAL   Cell 29 (2), 671-679 (1982)
  MEDLINE   83001943
FEATURES             Location/Qualifiers
     source          1..2590
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     conflict        28
                     /note="T is missing in [2]"
                     /citation=[2]
     conflict        80..82
                     /note="CGC is GCG in [2]"
                     /citation=[2]
     conflict        120
                     /note="C is T in [2]"
                     /citation=[2]
     conflict        135
                     /note="T is C in [2]"
                     /citation=[2]
     CDS             join(<215..509,901..951,1095..1139,1283..1327,
                     1471..1515,1634..1963,2061..>2380)
                     /codon_start=1
                     /product="C gamma 3"
                     /protein_id="CAA27268.1"
```

FIGURE 7

```
                     /db_xref="GI:577056"
/translation="ASTKGPSVFPLAPCSRSTSGGTAALGCLVKDYFPEPVTVSWNSG
ALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYTCNVNHKPSNTKVDKRVELKT
PLGDTTHTCPRCPEPKSCDTPPPCPRCPEPKSCDTPPPCPRCPEPKSCDTPPPCPRCP
APELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVQFKWYVDGVEVHNAK
TKPREEQYNSTFRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKTKGQPREP
QVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESSGQPENNYNTTPPMLDSDGSF
                     FLYSKLTVDKSRWQQGNIFSCSVMHEALHNRFTQKSLSLSPGK"
     intron          510..900
                     /note="intron I"
     intron          952..1094
                     /note="intron II"
     intron          1140..1282
                     /note="intron III"
     intron          1328..1470
                     /note="intron IV"
     intron          1516..1633
                     /note="intron V"
     intron          1964..2060
                     /note="intron VI"
     misc_feature    2484..2489
                     /note="pot. polyA signal"
BASE COUNT      541 a    925 c    703 g    421 t
ORIGIN
        1 agctttctgg ggcaggccag gcctgacttt ggctgggggc agggaggggg ctaaggtgac
       61 gcaggtggcg ccagccaggc gcacacccaa tgcccgtgag cccagacact ggaccctgcc
      121 tggaccctcg tggatagaca agaaccgagg ggcctctgcg ccctgggccc agctctgtcc
      181 cacaccgcag tcacatggcg ccatctctct tgcagcttcc accaagggcc catcggtctt
      241 ccccctggcg ccctgctcca ggagcacctc tgggggcaca gcggccctgg gctgcctggt
      301 caaggactac ttccccgaac cggtgacggt gtcgtggaac tcaggcgccc tgaccagcgg
      361 cgtgcacacc ttccccggctg tcctacagtc ctcaggactc tactccctca gcagcgtggt
      421 gaccgtgccc tccagcagct tgggcaccca gacctacacc tgcaacgtga atcacaagcc
      481 cagcaacacc aaggtggaca agagagttgg tgagaggcca gcgcagggag ggagggtgtc
      541 tgctggaagc caggctcagc cctcctgcct ggacgcatcc cggctgtgca gtcccagccc
      601 agggcagcaa ggcaggcccc gtctgactcc tcacccggag cctctgcccg ccccactcat
      661 gctcagggag agggtcttct ggctttttcc accaggctcc gggcaggcac aggctggatg
      721 cccctacccc aggcccttca cacacagggg caggtgctgc gctcagagct gccaaaagcc
      781 atatccagga ggaccctgcc cctgacctaa gcccacccca aaggccaaac tctctactca
      841 ctcagctcag acaccttctc tcttcccaga tctgagtaac tcccaatctt ctctctgcag
      901 agctcaaaac cccacttggt gacacaactc acacatgccc acggtgccca ggtaagccag
      961 cccaggactc gccctccagc tcaaggcggg acaagagccc tagagtggcc tgagtccagg
     1021 gacaggcccc agcagggtgc tgacgcatcc acctccatcc cagatccccg taactcccaa
     1081 tcttctctct gcagagccca aatcttgtga cacacctccc ccgtgcccac ggtgcccagg
     1141 taagccagcc caggcctcac cctccagctc aaggcaggac aagagcccta gagtggcctg
     1201 agtccaggga caggccccag cagggtgctg acgcgtccac ctccatccca gatccccgta
     1261 actcccaatc ttctctctgc agagcccaaa tcttgtgaca cacctccccc atgcccacgg
     1321 tgcccaggta agccagccca ggcctcgccc tccagctcaa ggcgggacaa gagccctaga
     1381 gtggcctgag tccagggaca ggccccagca gggtgctgac gcatccacct ccatcccaga
     1441 tccccgtaac tcccaatctt ctctctgcag agcccaaatc ttgtgacaca cctccccccgt
     1501 gcccaaggtg cccaggtaag ccagcccagg cctcgccctc cagctcaagg caggacaggt
     1561 gccctagagt ggcctgcatc cagggacagg tcccagtcgg gtgctgacac atctgcctcc
```

FIGURE 7 P

```
1621 atctcttcct cagcacctga actcctggga ggaccgtcag tcttcctctt ccccccaaaa
1681 cccaaggata cccttatgat ttcccggacc cctgaggtca cgtgcgtggt ggtggacgtg
1741 agccacgaag accccgaggt ccagttcaag tggtacgtgg acggcgtgga ggtgcataat
1801 gccaagacaa agccgcggga ggagcagtac aacagcacgt tccgtgtggt cagcgtcctc
1861 accgtcctgc accaggactg gctgaacggc aaggagtaca agtgcaaggt ctccaacaaa
1921 gccctcccag ccccccatcga gaaaaccatc tccaaaacca aaggtgggac ccgcggggta
1981 tgagggccac atggacagag gccagcttga cccaccctct gccctgggag tgaccgctgt
2041 gccaacctct gtccctacag gacagccccg agaaccacag gtgtacaccc tgcccccatc
2101 ccgggaggag atgaccaaga accaggtcag cctgacctgc ctggtcaaag gcttctaccc
2161 cagcgacatc gccgtggagt gggagagcag cgggcagccg gagaacaact acaacaccac
2221 gcctcccatg ctggactccg acggctcctt cttcctctac agcaagctca ccgtggacaa
2281 gagcaggtgg cagcagggga acatcttctc atgctccgtg atgcatgagg ctctgcacaa
2341 ccgcttcacg cagaagagcc tctccctgtc tccgggtaaa tgagtgcgac agccggcaag
2401 ccccgctcc ccgggctctc ggggtcgcgc gaggatgctt ggcacgtacc ccgtgtacat
2461 acttcccggg cacccagcat ggaaataaag cacccagcgc tgccctgggc ccctgtgaga
2521 ctgtgatggt tctttccacg ggtcaggccg agtctgaggc ctgagtgaca tgagggaggc
2581 agagcgggtc
```

VI. HUMAN IG GAMMA-4 CHAIN C REGION - HOMO SAPIENS (HUMAN).

AMINO ACID SEQUENCE

>sp|P01861|GC4_HUMAN IG GAMMA-4 CHAIN C REGION - Homo sapiens (Human).

```
        10         20         30         40         50         60
         |          |          |          |          |          |
ASTKGPSVFP LAPCSRSTSE STAALGCLVK DYFPEPVTVS WNSGALTSGV HTFPAVLQSS

        70         80         90        100        110        120
         |          |          |          |          |          |
GLYSLSSVVT VPSSSLGTKT YTCNVDHKPS NTKVDKRVES KYGPPCPSCP APEFLGGPSV

       130        140        150        160        170        180
         |          |          |          |          |          |
FLFPPKPKDT LMISRTPEVT CVVVDVSQED PEVQFNWYVD GVEVHNAKTK PREEQFNSTY

       190        200        210        220        230        240
         |          |          |          |          |          |
RVVSVLTVLH QDWLNGKEYK CKVSNKGLPS SIEKTISKAK GQPREPQVYT LPPSQEEMTK

       250        260        270        280        290        300
         |          |          |          |          |          |
NQVSLTCLVK GFYPSDIAVE WESNGQPENN YKTTPPVLDS DGSFFLYSRL TVDKSRWQEG

       310        320     327
         |          |       |
NVFSCSVMHE ALHNHYTQKS LSLSLGK
```

CODING SEQUENCE

```
                                                                 g       -1
cttccaccaa gggcccatcc gtcttccccc tggcgccctg ctccaggagc acctccgaga        60
gcacagccgc cctgggctgc ctggtcaagg actacttccc cgaaccggtg acggtgtcgt       120
ggaactcagg cgccctgacc agcggcgtgc acaccttccc ggctgtccta cagtcctcag       180
gactctactc cctcagcagc gtggtgaccg tgccctccag cagcttgggc acgaagacct       240
acacctgcaa cgtagatcac aagcccagca acaccaaggt ggacaagaga gttgagtcca       300
aatatggtcc cccatgccca tcatgcccag cacctgagtt cctgggggga ccatcagtct       360
tcctgttccc cccaaaaccc aaggacactc tcatgatctc ccggaccccc gaggtcacgt       420
```

FIGURE 7 Q

```
gcgtggtggt ggacgtgagc caggaagacc ccgaggtcca gttcaactgg tacgtggatg      480
gcgtggaggt gcataatgcc aagacaaagc cgcgggagga gcagttcaac agcacgtacc      540
gtgtggtcag cgtcctcacc gtcctgcacc aggactggct gaacggcaag gagtacaagt      600
gcaaggtctc caacaaaggc ctcccgtcct ccatcgagaa aaccatctcc aaagccaaag      660
ggcagccccg agagccacag gtgtacaccc tgcccccatc ccaggaggag atgaccaaga      720
accaggtcag cctgacctgc ctggtcaaag gcttctaccc cagcgacatc gccgtggagt      780
gggagagcaa tgggcagccg gagaacaact acaagaccac gcctcccgtg ctggactccg      840
acggctcctt cttcctctac agcaggctaa ccgtggacaa gagcaggtgg caggaggga       900
atgtcttctc atgctccgtg atgcatgagg ctctgcacaa ccactacaca cagaagagcc      960
tctccctgtc tctgggtaaa tga                                             983
```

GenBank

K01316. Human Ig germline

```
LOCUS       HUMIGCD2     2028 bp    DNA             PRI       11-APR-2001
DEFINITION  Human Ig germline H-chain G-E-A region B: gamma-4 constant region,
            3' end.
ACCESSION   K01316
VERSION     K01316.1  GI:184751
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2028)
  AUTHORS   Ellison,J., Buxbaum,J. and Hood,L.
  TITLE     Nucleotide sequence of a human immunoglobulin C gamma 4 gene
  JOURNAL   DNA 1 (1), 11-18 (1981)
  MEDLINE   83157104
REFERENCE   2  (bases 475 to 1069; 1180 to 1331; 1432 to 1655)
  AUTHORS   Takahashi,N., Ueda,S., Obata,M., Nikaido,T., Nakai,S. and Honjo,T.
  TITLE     Structure of human immunoglobulin gamma genes: implications for
            evolution of a gene family
  JOURNAL   Cell 29 (2), 671-679 (1982)
  MEDLINE   83001943
   PUBMED   6811139
REFERENCE   3  (bases 894 to 1106)
  AUTHORS   Krawinkel,U. and Rabbitts,T.H.
  TITLE     Comparison of the hinge-coding segments in human immunoglobulin
            gamma heavy chain genes and the linkage of the gamma 2 and gamma 4
            subclass genes
  JOURNAL   EMBO J. 1 (4), 403-407 (1982)
  MEDLINE   84235992
   PUBMED   6329676
REFERENCE   4  (bases 1 to 2028)
  AUTHORS   Ellison,J. and Hood,L.
  TITLE     Linkage and sequence homology of two human immunoglobulin gamma
            heavy chain constant region genes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 79 (6), 1984-1988 (1982)
  MEDLINE   82197621
   PUBMED   6804948
COMMENT     [1] reports that the human C-gamma-4 gene is equally homologous to
            the mouse gamma-1, gamma-2a, and gamma-2b genes (about 75%). [3]
            also reports partial sequences for human gamma-2, gamma-3, and a
            gamma pseudogene. [2] presents the gamma-1, gamma-2, gamma-3, and
            pseudo-gamma hinge regions.
            This entry is part of a multigene region (region B), which includes
            the gamma-2, gamma-4, epsilon-1, and alpha-2 genes. See segment 1
            for more comments.
            Complete source information:
            Human fetal liver DNA, library of T. Maniatis [3] and Lawn et al
            [1],[2]; clones 24B [1], lambda-HG4.1 [3], and Ig-gamma-4-2 [2].
FEATURES             Location/Qualifiers
     source          1..2028
```

FIGURE 7 R

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="14q32.33"
                     /germline
     gene            <1..2028
                     /gene="IgH"
                     /note="IGHG4"
     intron          <1..215
                     /gene="IgH"
                     /note="gamma-4 intron J-C"
     CDS             join(<216..509,900..935,1054..1383,1481..1803)
                     /gene="IgH"
                     /codon_start=3
                     /product="immunoglobulin gamma-4 heavy chain"
                     /protein_id="AAB59394.1"
                     /db_xref="GI:184759"
                     /translation="STKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGA
                     LTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTKTYTCNVDHKPSNTKVDKRVESKYG
                     PPCPSCPAPEFLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSQEDPEVQFNWYVDG
                     VEVHNAKTKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKGLPSSIEKTISKA
                     KGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPV
                     LDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALHNHYTQKSLSLSLGK"
     exon            216..509
                     /gene="IgH"
                     /note="G00-119-340"
     intron          510..899
                     /gene="IgH"
     exon            900..935
                     /gene="IgH"
     intron          936..1053
                     /gene="IgH"
     exon            1054..1383
                     /gene="IgH"
     intron          1384..1480
                     /gene="IgH"
     exon            1481..1803
                     /gene="IgH"
BASE COUNT      421 a    709 c    567 g    331 t
ORIGIN
        1 agctttctgg ggcaggccgg gcctgacttt ggctgggggc agggaggggg ctaaggtgac
       61 gcaggtggcg ccagccaggt gcacacccaa tgcccatgag cccagacact ggaccctgca
      121 tggaccatcg cggatagaca agaaccgagg ggcctctgcg ccctgggccc agctctgtcc
      181 cacaccgcgg tcacatggca ccacctctct tgcagcttcc accaagggcc catccgtctt
      241 ccccctggcg ccctgctcca ggagcacctc cgagagcaca gccgccctgg gctgcctggt
      301 caaggactac ttccccgaac cggtgacggt gtcgtggaac tcaggcgccc tgaccagcgg
      361 cgtgcacacc ttccccgctg tcctacagtc ctcaggactc tactccctca gcagcgtggt
      421 gaccgtgccc tccagcagct tgggcacgaa gacctacacc tgcaacgtag atcacaagcc
      481 cagcaacacc aaggtggaca agagagttgg tgagaggcca gcacagggag ggagggtgtc
      541 tgctggaagc caggctcagc cctcctgcct ggacgcaccc cggctgtgca gccccagccc
      601 agggcagcaa ggcatgcccc atctgtctcc tcacccggag gcctctgacc accccactca
      661 tgctcaggga gagggtcttc tggatttttc caccaggctc ccggcaccac aggctggatg
      721 cccctacccc aggccctgcg catacagggc aggtgctgcg ctcagacctg ccaagagcca
      781 tatccgggag gaccctgccc ctgacctaag cccaccccaa aggccaaact ctccactccc
      841 tcagctcaga cacttttctt cctcccagat ctgagtaact cccaatcttc tctctgcaga
      901 gtccaaatat ggtcccccat gcccatcatg cccaggtaag ccaacccagg cctcgccctc
      961 cagctcaagg cgggacaggt gccctagagt agcctgcatc cagggacagg ccccagccgg
     1021 gtgctgacgc atccacctcc atctcttcct cagcacctga gttcctgggg ggaccatcag
     1081 tcttcctgtt cccccaaaa cccaaggaca ctctcatgat ctcccggacc cctgaggtca
     1141 cgtgcgtggt ggtggacgtg agccaggaag accccgaggt ccagttcaac tggtacgtgg
     1201 atggcgtgga ggtgcataat gccaagacaa agccgcggga ggagcagttc aacagcacgt
     1261 accgtgtggt cagcgtcctc accgtcctgc accaggactg gctgaacggc aaggagtaca
     1321 agtgcaaggt ctccaacaaa ggcctcccgt cctccatcga gaaaaccatc tccaaagcca
     1381 aaggtgggac ccacggggtg cgagggccac acggacagag gccagctcgg cccaccctct
```

FIGURE 7 S

```
1441 gccctgggag tgaccgctgt gccaacctct gtccctacag ggcagccccg agagccacag
1501 gtgtacaccc tgcccccatc ccaggaggag atgaccaaga accaggtcag cctgacctgc
1561 ctggtcaaag gcttctaccc cagcgacatc gccgtggagt gggagagcaa tgggcagccg
1621 gagaacaact acaagaccac gcctcccgtg ctggactccg acggctcctt cttcctctac
1681 agcaggctaa ccgtggacaa gagcaggtgg caggaggggа atgtcttctc atgctccgtg
1741 atgcatgagg ctctgcacaa ccactacaca cagaagagcc tctccctgtc tctgggtaaa
1801 tgagtgccag ggccggcaag ccccgctccc ccgggctctc ggggtcgcgc gaggatgctt
1861 ggcacgtacc ccgtctacat acttcccagg cacccagcat ggaaataaag cacccaccac
1921 tgccctgggc ccctgtgaga ctgtgatggt tctttccacg ggtcaggccg agtctgaggc
1981 ctgagtgaca tgagggaggc agagcgggtc ccactgtccc cacactgg
```

VII. HUMAN IG DELTA CHAIN C REGION - HOMO SAPIENS (HUMAN).

AMINO ACID SEQUENCE

>sp|P01380|DTC_HUMAN IG DELTA CHAIN C REGION - Homo sapiens (Human).

```
        10         20         30         40         50         60
         |          |          |          |          |          |
APTKAPDVFP IISGCRHPKD NSPVVLACLI TGYHPISVTV TWYMGTQSQP QRTFPEIQRR

        70         80         90        100        110        120
         |          |          |          |          |          |
DSYYMTSSQL STPLQQWRQG EYKCVVQHTA SKSKKEIFRW PESPKAQASS VPTAQPQAEG

       130        140        150        160        170        180
         |          |          |          |          |          |
SLAKATTAPA TTRNTGRGGE EKKKEKEKEE QEERETKTPE CPSHTQPLGV YLLTPAVQDL

       190        200        210        220        230        240
         |          |          |          |          |          |
WLRDKATFTC FVVGSDLKDA HLTWEVAGKV PTGGVEEGLL ERHSNGSQSQ HSRLTLPRSL

       250        260        270        280        290        300
         |          |          |          |          |          |
WNAGTSVTCT LNHPSLPPQR LMALREPAAQ APVKLSLNLL ASSDPPEAAS WLLCEVSGFS

       310        320        330        340        350        360
         |          |          |          |          |          |
PPNILLMWLE DQREVNTSGF APARPPPQPG STTFWAWSVL RVPAPPSPQP ATYTCVVSHE

       370        380
         |          |
DSRTLLNASR SLEVSYVTDH GPM
```

GenBank

K02876. Human germline IgD...[gi:184766] PubMed, Protein, Related Sequences, Taxonomy, OMIM, LinkOut

```
LOCUS       HUMIGCH02     300 bp    DNA             PRI       08-NOV-1994
DEFINITION  Human germline IgD-chain gene, C-region, first hinge domain.
ACCESSION   K02876
VERSION     K02876.1  GI:184766
KEYWORDS    C-region; germline; hinge exon; immunoglobulin heavy chain;
            immunoglobulin-delta.
SOURCE      Homo sapiens (individual_isolate Chronic lymphocytic leukemia (CLL)
            patient) DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

FIGURE 7 T

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   White,M.B., Shen,A.L., Word,C.J., Tucker,P.W. and Blattner,F.R.
  TITLE     Human immunoglobulin D: genomic sequence of the delta heavy chain
  JOURNAL   Science 228 (4700), 733-737 (1985)
  MEDLINE   85192522
COMMENT     See segment 1.
FEATURES             Location/Qualifiers
     source          1..300
                     /organism="Homo sapiens"
                     /isolate="Chronic lymphocytic leukemia (CLL) patient"
                     /db_xref="taxon:9606"
                     /map="14q32.33"
                     /cell_type="lymphocyte"
                     /germline
     intron          <1..151
                     /gene="IGHD"
                     /note="G00-120-084"
                     /number=1
     exon            101..202
                     /partial
                     /gene="IGHD"
                     /note="hinge-1 domain; G00-120-084"
                     /number=2
     intron          203..>300
                     /gene="IGHD"
                     /note="G00-120-084"
                     /number=2
     gene            join(K02875.1:1..495,1..300,K02877.1:1..300,
                     K02878.1:1..500,K02879.1:1..500,K02880.1:1..100,
                     K02881.1:1..200,K02882.1:1..52)
                     /gene="IGHD"
     CDS             join(K02875.1:101..403,101..202,K02877.1:101..172,
                     K02878.1:101..424,K02879.1:101..424,K02881.1:25..182,
                     K02882.1:44..52)
                     /partial
                     /gene="IGHD"
                     /note="membrane bound form"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAA52771.1"
                     /db_xref="GI:495872"
                     /db_xref="GDB:G00-120-084"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLACLITGYHPTSVTVTWYMG
                     TQSQPQRTFPEIQRRDSYYMTSSQLSTPLQQWRQGEYKCVVQHTASKSKKEIFRWPES
                     PKAQASSVPTAQPQAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEEQEERETKTPEC
                     PSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGL
                     LERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPPQRLMALREPAAQAPVKLSL
                     NLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPRSTTFW
                     AWSVLRVPAPPSPQPATYTCVVSHEDSRTLLNASRSLEVSYLAMTPLIPQSKDENSDD
                     YTTFDDVGSLWTTLSTFVALFILTLLYSGIVTFIKVK"
     CDS             join(K02875.1:101..403,101..202,K02877.1:101..172,
                     K02878.1:101..424,K02879.1:101..424,K02880.1:25..53)
                     /partial
                     /gene="IGHD"
                     /note="secreted form"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAA52770.1"
                     /db_xref="GI:495871"
                     /db_xref="GDB:G00-120-084"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLACLITGYHPTSVTVTWYMG
                     TQSQPQRTFPEIQRRDSYYMTSSQLSTPLQQWRQGEYKCVVQHTASKSKKEIFRWPES
```

FIGURE 7 U

```
                    PKAQASSVPTAQPQAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEEQEERETKTPEC
                    PSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGL
                    LERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPPQRLMALREPAAQAPVKLSL
                    NLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPRSTTFW
                    AWSVLRVPAPPSPQPATYTCVVSHEDSRTLLNASRSLEVSYVTDHGPMK"
BASE COUNT       59 a    133 c     52 g     56 t
ORIGIN      About 300 bp after segment 1; 118 bp upstream of StuI site.
        1 taggctgcct gtgcccccca cctgcctgtc cacaacccag cctctggtac atccatgccc
       61 tctgccctaa gcctcacctg cacttttcct tggatttcag agtctccaaa ggcacaggcc
      121 tcctccgtgc ccactgcaca accccaagca gagggcagcc tcgccaaggc aaccacagcc
      181 ccagccacca cccgtaacac aggtgagaag ccccttcccc gcacactcca cccccaccca
      241 cctgctcatt cctcagccgc ctcctccagg cagcccttca taactccttg tctgagtctc

K02877. Human Ig germline ...[gi:184767] PubMed, Protein, Related Sequences, Taxonomy,
OMIM, LinkOut

LOCUS       HUMIGCH03     300 bp    DNA             PRI       08-NOV-1994
DEFINITION  Human Ig germline delta H-chain C-region gene, second hinge domain
            (CLL lymphocyte).
ACCESSION   K02877
VERSION     K02877.1  GI:184767
KEYWORDS    C-region; germline; hinge exon; immunoglobulin heavy chain;
            immunoglobulin-delta.
SOURCE      Homo sapiens (individual_isolate Chronic lymphocytic leukemia (CLL)
            patient) DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   White,M.B., Shen,A.L., Word,C.J., Tucker,P.W. and Blattner,F.R.
  TITLE     Human immunoglobulin D: genomic sequence of the delta heavy chain
  JOURNAL   Science 228 (4700), 733-737 (1985)
  MEDLINE   85192522
COMMENT     See segment 1.
FEATURES             Location/Qualifiers
     source          1..300
                     /organism="Homo sapiens"
                     /isolate="Chronic lymphocytic leukemia (CLL) patient"
                     /db_xref="taxon:9606"
                     /map="14q32.33"
                     /cell_type="lymphocyte"
                     /germline
     intron          <1..100
                     /gene="IGHD"
                     /note="G00-120-084"
                     /number=2
     exon            101..172
                     /gene="IGHD"
                     /note="hinge-2 domain; G00-120-084; putative"
                     /number=3
     intron          173..>300
                     /gene="IGHD"
                     /note="G00-120-084"
                     /number=3
     gene            join(K02875.1:1..495,K02876.1:1..300,1..300,
                     K02878.1:1..500,K02879.1:1..500,K02880.1:1..100,
                     K02881.1:1..200,K02882.1:1..52)
                     /gene="IGHD"
     CDS             join(K02875.1:101..403,K02876.1:101..202,101..172,
                     K02878.1:101..424,K02879.1:101..424,K02881.1:25..182,
                     K02882.1:44..52)
                     /partial
                     /gene="IGHD"
```

FIGURE 7 V

```
                     /note="membrane bound form"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAA52771.1"
                     /db_xref="GI:495872"
                     /db_xref="GDB:G00-120-084"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLACLITGYHPTSVTVTWYMG
                     TQSQPQRTFPEIQRRDSYYMTSSQLSTPLQQWRQGEYKCVVQHTASKSKKEIFRWPES
                     PKAQASSVPTAQPQAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEEQEERETKTPEC
                     PSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGL
                     LERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPPQRLMALREPAAQAPVKLSL
                     NLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPRSTTFW
                     AWSVLRVPAPPSPQPATYTCVVSHEDSRTLLNASRSLEVSYLAMTPLIPQSKDENSDD
                     YTTFDDVGSLWTTLSTFVALFILTLLYSGIVTFIKVK"
     CDS             join(K02875.1:101..403,K02876.1:101..202,101..172,
                     K02878.1:101..424,K02879.1:101..424,K02880.1:25..53)
                     /partial
                     /gene="IGHD"
                     /note="secreted form"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAA52770.1"
                     /db_xref="GI:495871"
                     /db_xref="GDB:G00-120-084"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLACLITGYHPTSVTVTWYMG
                     TQSQPQRTFPEIQRRDSYYMTSSQLSTPLQQWRQGEYKCVVQHTASKSKKEIFRWPES
                     PKAQASSVPTAQPQAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEEQEERETKTPEC
                     PSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGL
                     LERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPPQRLMALREPAAQAPVKLSL
                     NLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPRSTTFW
                     AWSVLRVPAPPSPQPATYTCVVSHEDSRTLLNASRSLEVSYVTDHGPMK"
BASE COUNT      102 a     52 c     70 g     76 t
ORIGIN      About 1.85 kb after segment 2.
        1 gtcattagct ggatttagcc attccacaat gtacacatat ttcaaacatt gtgttgtata
       61 tgataaacat gtataatttt tgtcaattaa aaatttttag gaagaggagg agaagagaag
      121 aagaaggaga aggagaaaga ggaacaagaa gagagagaga caaagacacc aggttttttc
      181 tgacccctgg gctatcaaaa cacctattgc ccaataacta gttggccgtt ggtgccctaa
      241 actattgaag cgattgctgt tatgtggatg ggccccggac acttagaaac tcgtgacccc

K02878. Human germline IgD...[gi:184768] PubMed, Protein, Related Sequences, Taxonomy,
OMIM, LinkOut

LOCUS       HUMIGCH04      500 bp    DNA             PRI       08-NOV-1994
DEFINITION  Human germline IgD chain gene, C-region, C-delta-2 domain.
ACCESSION   K02878
VERSION     K02878.1  GI:184768
KEYWORDS    C-region; germline; immunoglobulin heavy chain;
            immunoglobulin-delta.
SOURCE      Homo sapiens (individual_isolate Chronic lymphocytic leukemia (CLL)
            patient) DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   White,M.B., Shen,A.L., Word,C.J., Tucker,P.W. and Blattner,F.R.
  TITLE     Human immunoglobulin D: genomic sequence of the delta heavy chain
  JOURNAL   Science 228 (4700), 733-737 (1985)
  MEDLINE   85192522
COMMENT     See segment 1.
FEATURES             Location/Qualifiers
     source          1..500
                     /organism="Homo sapiens"
                     /isolate="Chronic lymphocytic leukemia (CLL) patient"
```

FIGURE 7

```
                     /db_xref="taxon:9606"
                     /map="14q32.33"
                     /cell_type="lymphocyte"
                     /germline
     intron          <1..100
                     /gene="IGHD"
                     /note="G00-120-084"
                     /number=3
     exon            101..424
                     /gene="IGHD"
                     /note="C-delta-2 domain; G00-120-084"
                     /number=4
     intron          425..>500
                     /gene="IGHD"
                     /note="G00-120-084"
                     /number=4
     intron          425..>500
                     /gene="IGHD"
                     /note="IgD-s intron D"
     gene            join(K02875.1:1..495,K02876.1:1..300,K02877.1:1..300,
                     1..500,K02879.1:1..500,K02880.1:1..100,K02881.1:1..200,
                     K02882.1:1..52)
                     /gene="IGHD"
     CDS             join(K02875.1:101..403,K02876.1:101..202,
                     K02877.1:101..172,101..424,K02879.1:101..424,
                     K02881.1:25..182,K02882.1:44..52)
                     /partial
                     /gene="IGHD"
                     /note="membrane bound form"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAA52771.1"
                     /db_xref="GI:495872"
                     /db_xref="GDB:G00-120-084"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLACLITGYHPTSVTVTWYMG
                     TQSQPQRTFPEIQRRDSYYMTSSQLSTPLQQWRQGEYKCVVQHTASKSKKEIFRWPES
                     PKAQASSVPTAQPQAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEEQEERETKTPEC
                     PSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGL
                     LERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPPQRLMALREPAAQAPVKLSL
                     NLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPRSTTFW
                     AWSVLRVPAPPSPQPATYTCVVSHEDSRTLLNASRSLEVSYLAMTPLIPQSKDENSDD
                     YTTFDDVGSLWTTLSTFVALFILTLLYSGIVTFIKVK"
     CDS             join(K02875.1:101..403,K02876.1:101..202,
                     K02877.1:101..172,101..424,K02879.1:101..424,
                     K02880.1:25..53)
                     /partial
                     /gene="IGHD"
                     /note="secreted form"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAA52770.1"
                     /db_xref="GI:495871"
                     /db_xref="GDB:G00-120-084"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLACLITGYHPTSVTVTWYMG
                     TQSQPQRTFPEIQRRDSYYMTSSQLSTPLQQWRQGEYKCVVQHTASKSKKEIFRWPES
                     PKAQASSVPTAQPQAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEEQEERETKTPEC
                     PSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGL
                     LERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPPQRLMALREPAAQAPVKLSL
                     NLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPRSTTFW
                     AWSVLRVPAPPSPQPATYTCVVSHEDSRTLLNASRSLEVSYVTDHGPMK"
BASE COUNT       93 a    171 c    157 g     79 t
ORIGIN      About 450 bp after segment 3; 131 bp upstream of AccI site.
        1 gaagctgggg agaggagagc acagtggtta agtcagtccc tgcagcccaa ctgctcccga
```

FIGURE 7 X

```
      61 aggtccggcc acagctgctc tcgtttgctc tcccctgcag agtgtccgag ccacacccag
     121 cctcttggcg tctacctgct aacccctgca gtgcaggacc tgtggctccg ggacaaagcc
     181 accttcacct gcttcgtggt gggcagtgac ctgaaggatg ctcacctgac ctgggaggtg
     241 gctgggaagg tccccacagg gggcgtggag gaagggctgc tggagcggca cagcaacggc
     301 tcccagagcc agcacagccg tctgaccctg cccaggtcct tgtggaacgc ggggacctcc
     361 gtcacctgca cactgaacca tcccagcctc ccaccccaga ggttgatggc gctgagagaa
     421 cccggtgagc ctggctccca ggtggggaga cgagggtgcc cacagcctgc tgaccctac
     481 gcccgcccca gggccatgac
```

K02879. Human Ig germline ...[gi:184769] PubMed, Protein, Related Sequences, Taxonomy, OMIM, LinkOut

```
LOCUS       HUMIGCH05      500 bp    DNA             PRI       08-NOV-1994
DEFINITION  Human Ig germline delta H-chain C-region gene, C-delta-3 domain
            (CLL lymphocyte).
ACCESSION   K02879
VERSION     K02879.1  GI:184769
KEYWORDS    C-region; germline; immunoglobulin heavy chain;
            immunoglobulin-delta.
SOURCE      Homo sapiens (individual_isolate Chronic lymphocytic leukemia (CLL)
            patient) DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   White,M.B., Shen,A.L., Word,C.J., Tucker,P.W. and Blattner,F.R.
  TITLE     Human immunoglobulin D: genomic sequence of the delta heavy chain
  JOURNAL   Science 228 (4700), 733-737 (1985)
  MEDLINE   85192522
COMMENT     See segment 1.
FEATURES             Location/Qualifiers
     source          1..500
                     /organism="Homo sapiens"
                     /isolate="Chronic lymphocytic leukemia (CLL) patient"
                     /db_xref="taxon:9606"
                     /map="14q32.33"
                     /cell_type="lymphocyte"
                     /germline
     intron          <1..100
                     /gene="IGHD"
                     /note="G00-120-084"
                     /number=4
     exon            101..424
                     /gene="IGHD"
                     /note="C-delta-3 domain; G00-120-084; putative"
                     /number=5
     intron          425..>500
                     /gene="IGHD"
                     /note="G00-120-084"
                     /number=5
     gene            join(K02875.1:1..495,K02876.1:1..300,K02877.1:1..300,
                     K02878.1:1..500,1..500,K02880.1:1..100,K02881.1:1..200,
                     K02882.1:1..52)
                     /gene="IGHD"
     CDS             join(K02875.1:101..403,K02876.1:101..202,
                     K02877.1:101..172,K02878.1:101..424,101..424,
                     K02881.1:25..182,K02882.1:44..52)
                     /partial
                     /gene="IGHD"
                     /note="membrane bound form"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAA52771.1"
```

FIGURE 7 Y

```
                     /db_xref="GI:495872"
                     /db_xref="GDB:G00-120-084"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLACLITGYHPTSVTVTWYMG
                     TQSQPQRTFPEIQRRDSYYMTSSQLSTPLQQWRQGEYKCVVQHTASKSKKEIFRWPES
                     PKAQASSVPTAQPQAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEEQEERETKTPEC
                     PSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGL
                     LERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPPQRLMALREPAAQAPVKLSL
                     NLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPRSTTFW
                     AWSVLRVPAPPSPQPATYTCVVSHEDSRTLLNASRSLEVSYLAMTPLIPQSKDENSDD
                     YTTFDDVGSLWTTLSTFVALFILTLLYSGIVTFIKVK"
     CDS             join(K02875.1:101..403,K02876.1:101..202,
                     K02877.1:101..172,K02878.1:101..424,101..424,
                     K02880.1:25..53)
                     /partial
                     /gene="IGHD"
                     /note="secreted form"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAA52770.1"
                     /db_xref="GI:495871"
                     /db_xref="GDB:G00-120-084"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLACLITGYHPTSVTVTWYMG
                     TQSQPQRTFPEIQRRDSYYMTSSQLSTPLQQWRQGEYKCVVQHTASKSKKEIFRWPES
                     PKAQASSVPTAQPQAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEEQEERETKTPEC
                     PSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGL
                     LERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPPQRLMALREPAAQAPVKLSL
                     NLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPRSTTFW
                     AWSVLRVPAPPSPQPATYTCVVSHEDSRTLLNASRSLEVSYVTDHGPMK"
BASE COUNT       85 a    188 c    145 g     82 t
ORIGIN      About 150 bp after segment 4; 118 bp upstream of HindIII site.
        1 ccacaggaaa ggagaaggga ggcaccacac cctggccggc cccacttctc tcccagtgcc
       61 cccgtggcca gagcctgaca gccccccccac ctccccgcag ctgcgcaggc acccgtcaag
      121 ctttctctga acctgctggc ctcgtctgac cctcccgagg cggcctcgtg gctcctgtgt
      181 gaggtgtctg gcttctcgcc ccccaacatc ctcctgatgt ggctggagga ccagcgtgag
      241 gtgaacactt ctgggtttgc ccccgcacgc ccccctccac agcccaggag caccacgttc
      301 tgggcctgga gtgtgctgcg tgtcccagcc ccgcccagcc ctcagccagc cacctacacg
      361 tgtgtggtca gccacgagga ctcccggact ctgctcaacg ccagccggag cctagaagtc
      421 agctgtgagt caccccagg ccagggttgg gacggggact ctgaggggg ccataaggag
      481 ctggaatcca tactaggcag

K01311. Human IgD germline...[gi:184716] PubMed, Protein, Taxonomy, OMIM

LOCUS       HUMIGCB9      106 bp    DNA             PRI       12-APR-2001
DEFINITION  Human IgD germline chain J-delta region: C-delta CH1.
ACCESSION   K01311
VERSION     K01311.1  GI:184716
KEYWORDS    C-region; germline; immunoglobulin heavy chain;
            immunoglobulin-delta.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 106)
  AUTHORS   Rabbitts,T.H., Forster,A. and Milstein,C.P.
  TITLE     Human immunoglobulin heavy chain genes: evolutionary comparisons of
            C mu, C delta and C gamma genes and associated switch sequences
  JOURNAL   Nucleic Acids Res. 9 (18), 4509-4524 (1981)
  MEDLINE   82059479
   PUBMED   6795593
COMMENT     The deduced amino acid sequence is compared in [1] to the
            J/C-delta-1 junction of human ER1 protein. The delta gene occurs
            only 5 kb from the mu region. The authors [1] could not detect any
            switch-related sequences adjacent to the delta gene and state that
```

FIGURE 7

```
            this implies that the mu/delta switch cannot occur by the class
            switch recombination method. They speculate that the entire
            VH-(C-mu)-(C-delta) region is transcribed into one nuclear
            precursor molecule which is spliced later.
            This is part of a multigene region containing the J-region, switch
            region, C-mu-secreted, C-mu-membrane, and C-delta genes.
FEATURES             Location/Qualifiers
     source          1..106
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="14q32.33"
                     /cell_type="lymphocyte"
                     /tissue_type="placenta"
                     /tissue_type="liver"
                     /dev_stage="foetus"
                     /germline
                     /tissue_lib="of Lawn et al."
     gene            1..106
                     /gene="IGHD"
     intron          <1..26
                     /gene="IGHD"
                     /note="intron delta  J-C; G00-120-084"
     CDS             <27..>106
                     /gene="IGHD"
                     /note="C-region CH1 domain"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAB59423.1"
                     /db_xref="GI:184735"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLA"
BASE COUNT       24 a     38 c     24 g     20 t
ORIGIN
        1 tgccacccca ggactctgtc ttccagcacc caccaaggct ccggatgtgt tccccatcat
       61 atcagggtgc agacacccaa aggataacag ccctgtggtc ctggca

K02880. Human germline IgD...[gi:184770] PubMed, Protein, Related Sequences, Taxonomy,
OMIM, LinkOut

LOCUS       HUMIGCH06     100 bp    DNA             PRI       08-NOV-1994
DEFINITION  Human germline IgD chain gene, C-region, secreted terminus.
ACCESSION   K02880
VERSION     K02880.1  GI:184770
KEYWORDS    C-region; germline; immunoglobulin heavy chain;
            immunoglobulin-delta.
SOURCE      Homo sapiens (individual_isolate Chronic lymphocytic leukemia (CLL)
            patient) DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100)
  AUTHORS   White,M.B., Shen,A.L., Word,C.J., Tucker,P.W. and Blattner,F.R.
  TITLE     Human immunoglobulin D: genomic sequence of the delta heavy chain
  JOURNAL   Science 228 (4700), 733-737 (1985)
  MEDLINE   85192522
COMMENT     See segment 1.
FEATURES             Location/Qualifiers
     source          1..100
                     /organism="Homo sapiens"
                     /isolate="Chronic lymphocytic leukemia (CLL) patient"
                     /db_xref="taxon:9606"
                     /map="14q32.33"
                     /cell_type="lymphocyte"
                     /germline
```

FIGURE 7 AA

```
     intron          <1..>100
                     /gene="IGHD"
                     /note="G00-120-084"
                     /number=5
     intron          <1..24
                     /gene="IGHD"
                     /note="G00-120-084"
                     /number=5
     exon            25..>53
                     /gene="IGHD"
                     /note="secreted terminus domain; G00-120-084"
                     /number=6
     gene            join(K02875.1:1..495,K02876.1:1..300,K02877.1:1..300,
                     K02878.1:1..500,K02879.1:1..500,1..100,K02881.1:1..200,
                     K02882.1:1..52)
                     /gene="IGHD"
     CDS             join(K02875.1:101..403,K02876.1:101..202,
                     K02877.1:101..172,K02878.1:101..424,K02879.1:101..424,
                     25..53)
                     /partial
                     /gene="IGHD"
                     /note="secreted form"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAA52770.1"
                     /db_xref="GI:495871"
                     /db_xref="GDB:G00-120-084"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLACLITGYHPTSVTVTWYMG
                     TQSQPQRTFPEIQRRDSYYMTSSQLSTPLQQWRQGEYKCVVQHTASKSKKEIFRWPES
                     PKAQASSVPTAQPQAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEEQEERETKTPEC
                     PSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGL
                     LERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPPQRLMALREPAAQAPVKLSL
                     NLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPRSTTFW
                     AWSVLRVPAPPSPQPATYTCVVSHEDSRTLLNASRSLEVSYVTDHGPMK"
BASE COUNT       24 a     33 c     22 g     21 t
ORIGIN      About 1.8 kb after segment 5.
        1 gacacgccga ttttttgtta ttagatgtaa cagaccatgg ccccatgaaa tgatcccgga
       61 ccagatccgt ccgcacccgc cactcagcag ctctggccga

K02881. Human germline IgD...[gi:184771] PubMed, Protein, Related Sequences, Taxonomy,
OMIM, LinkOut

LOCUS       HUMIGCH07     200 bp    DNA             PRI       08-NOV-1994
DEFINITION  Human germline IgD-chain gene, C-region, first domain of membrane
            terminus.
ACCESSION   K02881
VERSION     K02881.1  GI:184771
KEYWORDS    C-region; germline; immunoglobulin heavy chain;
            immunoglobulin-delta.
SOURCE      Homo sapiens (individual_isolate Chronic lymphocytic leukemia (CLL)
            patient) DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200)
  AUTHORS   White,M.B., Shen,A.L., Word,C.J., Tucker,P.W. and Blattner,F.R.
  TITLE     Human immunoglobulin D: genomic sequence of the delta heavy chain
  JOURNAL   Science 228 (4700), 733-737 (1985)
  MEDLINE   85192522
COMMENT     See segment 1.
FEATURES             Location/Qualifiers
     source          1..200
                     /organism="Homo sapiens"
```

FIGURE 7 BB

```
                     /isolate="Chronic lymphocytic leukemia (CLL) patient"
                     /db_xref="taxon:9606"
                     /map="14q32.33"
                     /cell_type="lymphocyte"
                     /germline
     intron          <1..24
                     /gene="IGHD"
                     /note="G00-120-084"
                     /number=5
     exon            25..182
                     /gene="IGHD"
                     /note="first domain of membrane terminus; G00-120-084;
                     putative"
                     /number=6
     intron          183..>200
                     /gene="IGHD"
                     /note="G00-120-084"
                     /number=6
     gene            join(K02875.1:1..495,K02876.1:1..300,K02877.1:1..300,
                     K02878.1:1..500,K02879.1:1..500,K02880.1:1..100,1..200,
                     K02882.1:1..52)
                     /gene="IGHD"
     CDS             join(K02875.1:101..403,K02876.1:101..202,
                     K02877.1:101..172,K02878.1:101..424,K02879.1:101..424,
                     25..182,K02882.1:44..52)
                     /partial
                     /gene="IGHD"
                     /note="membrane bound form"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAA52771.1"
                     /db_xref="GI:495872"
                     /db_xref="GDB:G00-120-084"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLACLITGYHPTSVTVTWYMG
                     TQSQPQRTFPEIQRRDSYYMTSSQLSTPLQQWRQGEYKCVVQHTASKSKKEIFRWPES
                     PKAQASSVPTAQPQAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEEQEERETKTPEC
                     PSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGL
                     LERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPPQRLMALREPAAQAPVKLSL
                     NLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPRSTTFW
                     AWSVLRVPAPPSPQPATYTCVVSHEDSRTLLNASRSLEVSYLAMTPLIPQSKDENSDD
                     YTTFDDVGSLWTTLSTFVALFILTLLYSGIVTFIKVK"
BASE COUNT       37 a     72 c     49 g     42 t
ORIGIN      About 800 bp after segment 6.
        1 cgctcggccc ccgttcctcc ccagacctgg ccatgacccc cctgatccct cagagcaagg
       61 atgagaacag cgatgactac acgacctttg atgatgtggg cagcctgtgg accaccctgt
      121 ccacgtttgt ggccctcttc atcctcaccc tcctctacag cggcattgtc actttcatca
      181 aggtcagggg agcggccagg

K02882. Human germline IgD...[gi:184772] PubMed, Protein, Related Sequences, Taxonomy,
OMIM, LinkOut

LOCUS       HUMIGCH08     100 bp    DNA             PRI       08-NOV-1994
DEFINITION  Human germline IgD-chain gene, C-region, second domain of membrane
            terminus.
ACCESSION   K02882
VERSION     K02882.1  GI:184772
KEYWORDS    C-region; germline; immunoglobulin heavy chain;
            immunoglobulin-delta.
SOURCE      Homo sapiens (individual_isolate Chronic lymphocytic leukemia (CLL)
            patient) DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

FIGURE 7 CC

```
REFERENCE   1  (bases 1 to 100)
  AUTHORS   White,M.B., Shen,A.L., Word,C.J., Tucker,P.W. and Blattner,F.R.
  TITLE     Human immunoglobulin D: genomic sequence of the delta heavy chain
  JOURNAL   Science 228 (4700), 733-737 (1985)
  MEDLINE   85192522
COMMENT     See segment 1.
FEATURES             Location/Qualifiers
     source          1..100
                     /organism="Homo sapiens"
                     /isolate="Chronic lymphocytic leukemia (CLL) patient"
                     /db_xref="taxon:9606"
                     /map="14q32.33"
                     /cell_type="lymphocyte"
                     /germline
     intron          <1..43
                     /gene="IGHD"
                     /note="IgD-Mb"
                     /number=6
     exon            44..>52
                     /gene="IGHD"
                     /note="membrane-bound form  (second domain of membrane
                     terminus); G00-120-084; putative"
                     /number=7
     gene            join(K02875.1:1..495,K02876.1:1..300,K02877.1:1..300,
                     K02878.1:1..500,K02879.1:1..500,K02880.1:1..100,
                     K02881.1:1..200,1..52)
                     /gene="IGHD"
     CDS             join(K02875.1:101..403,K02876.1:101..202,
                     K02877.1:101..172,K02878.1:101..424,K02879.1:101..424,
                     K02881.1:25..182,44..52)
                     /partial
                     /gene="IGHD"
                     /note="membrane bound form"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAA52771.1"
                     /db_xref="GI:495872"
                     /db_xref="GDB:G00-120-084"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLACLITGYHPTSVTVTWYMG
                     TQSQPQRTFPEIQRRDSYYMTSSQLSTPLQQWRQGEYKCVVQHTASKSKKEIFRWPES
                     PKAQASSVPTAQPQAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEEQEERETKTPEC
                     PSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGL
                     LERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPPQRLMALREPAAQAPVKLSL
                     NLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPRSTTFW
                     AWSVLRVPAPPSPQPATYTCVVSHEDSRTLLNASRSLEVSYLAMTPLIPQSKDENSDD
                     YTTFDDVGSLWTTLSTFVALFILTLLYSGIVTFIKVK"
BASE COUNT       22 a     30 c     30 g     18 t
ORIGIN      About 1.3 kb after segment 7.
        1 tcaggcttct agcccctgtc tgaccccagg ggctgtcttt caggtgaagt agccccagaa
       61 gagcaggacg ccctgtacct gcagagaagg gaagcagcct

K02875. Human germline IgD...[gi:184765] PubMed, Related Sequences, Taxonomy, OMIM,
LinkOut

LOCUS       HUMIGCH01     495 bp    DNA             PRI       08-NOV-1994
DEFINITION  Human germline IgD chain gene, C-region, C-delta-1 domain.
ACCESSION   K02875
VERSION     K02875.1  GI:184765
KEYWORDS    C-region; germline; immunoglobulin heavy chain;
            immunoglobulin-delta.
SOURCE      Homo sapiens (individual_isolate Chronic lymphocytic leukemia (CLL)
            patient) DNA.
  ORGANISM  Homo sapiens
```

FIGURE 7 DD

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   White,M.B., Shen,A.L., Word,C.J., Tucker,P.W. and Blattner,F.R.
  TITLE     Human immunoglobulin D: genomic sequence of the delta heavy chain
  JOURNAL   Science 228 (4700), 733-737 (1985)
  MEDLINE   85192522
COMMENT     Sequence in computer readable form and draft entry for [1] were
            kindly provided by M.B.White, 06-AUG-1985.
            The C-delta and delta-s exon boundaries were located by comparing
            the translated sequences with known AA sequences [1].
FEATURES             Location/Qualifiers
     source          1..495
                     /organism="Homo sapiens"
                     /isolate="Chronic lymphocytic leukemia (CLL) patient"
                     /db_xref="taxon:9606"
                     /map="14q32.33"
                     /cell_type="lymphocyte"
                     /germline
     gene            join(1..495,K02876.1:1..300,K02877.1:1..300,
                     K02878.1:1..500,K02879.1:1..500,K02880.1:1..100,
                     K02881.1:1..200,K02882.1:1..52)
                     /gene="IGHD"
     intron          <1..100
                     /gene="IGHD"
                     /note="J-C intron; G00-120-084"
     CDS             join(101..403,K02876.1:101..202,K02877.1:101..172,
                     K02878.1:101..424,K02879.1:101..424,K02880.1:25..53)
                     /partial
                     /gene="IGHD"
                     /note="secreted form"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAA52770.1"
                     /db_xref="GI:495871"
                     /db_xref="GDB:G00-120-084"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLACLITGYHPTSVTVTWYMG
                     TQSQPQRTFPEIQRRDSYYMTSSQLSTPLQQWRQGEYKCVVQHTASKSKKEIFRWPES
                     PKAQASSVPTAQPQAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEEQEERETKTPEC
                     PSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGL
                     LERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPPQRLMALREPAAQAPVKLSL
                     NLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPRSTTFW
                     AWSVLRVPAPPSPQPATYTCVVSHEDSRTLLNASRSLEVSYVTDHGPMK"
     exon            101..403
                     /gene="IGHD"
                     /note="C-delta-1 domain; G00-120-084; putative"
                     /number=1
     CDS             join(101..403,K02876.1:101..202,K02877.1:101..172,
                     K02878.1:101..424,K02879.1:101..424,K02881.1:25..182,
                     K02882.1:44..52)
                     /partial
                     /gene="IGHD"
                     /note="membrane bound form"
                     /codon_start=3
                     /product="immunoglobulin delta-chain"
                     /protein_id="AAA52771.1"
                     /db_xref="GI:495872"
                     /db_xref="GDB:G00-120-084"
                     /translation="PTKAPDVFPIISGCRHPKDNSPVVLACLITGYHPTSVTVTWYMG
                     TQSQPQRTFPEIQRRDSYYMTSSQLSTPLQQWRQGEYKCVVQHTASKSKKEIFRWPES
                     PKAQASSVPTAQPQAEGSLAKATTAPATTRNTGRGGEEKKKEKEKEEQEERETKTPEC
                     PSHTQPLGVYLLTPAVQDLWLRDKATFTCFVVGSDLKDAHLTWEVAGKVPTGGVEEGL
                     LERHSNGSQSQHSRLTLPRSLWNAGTSVTCTLNHPSLPPQRLMALREPAAQAPVKLSL
```

FIGURE 7 EE

```
                     NLLASSDPPEAASWLLCEVSGFSPPNILLMWLEDQREVNTSGFAPARPPPQPRSTTFW
                     AWSVLRVPAPPSPQPATYTCVVSHEDSRTLLNASRSLEVSYLAMTPLIPQSKDENSDD
                     YTTFDDVGSLWTTLSTFVALFILTLLYSGIVTFIKVK"
     intron          404..>495
                     /gene="IGHD"
                     /note="G00-120-084"
                     /number=1
BASE COUNT      114 a    179 c    120 g     82 t
ORIGIN      182 bp upstream of SphI site; chromosome 14q32.3.
        1 tttccctgcc tcccgtcacc ctgccgccag ggcctctgcc ctgccctgcc ccttgtcctc
       61 aggtttccag cctcagactc ccactgtgtc tgtcttccag cacccaccaa ggctccggat
      121 gtgttcccca tcatatcagg gtgcagacac ccaaaggata acagccctgt ggtcctggca
      181 tgcttgataa ctgggtacca cccaacgtcc gtgactgtca cctggtacat ggggacacag
      241 agccagcccc agagaacctt ccctgagata caaagacggg acagctacta catgacaagc
      301 agccagctct ccacccccct ccagcagtgg cgccaaggcg agtacaaatg cgtggtccag
      361 cacaccgcca gcaagagtaa gaaggagatc ttccgctggc caggtaggtc gcaccggaga
      421 tcacccagaa gggccccca ggacccccag caccttccac tcagggcctg accacaaaga
      481 cagaagcaag ggctg
```

VIII. HUMAN IG EPSILON CHAIN C REGION - HOMO SAPIENS (HUMAN).

AMINO ACID SEQUENCE

>sp|P01854|EPC_HUMAN IG EPSILON CHAIN C REGION - Homo sapiens (Human).

```
         10          20          30          40          50          60
          |           |           |           |           |           |
ASTQSPSVFP LTRCCKNIPS NATSVTLGCL ATGYFPEPVM VTWDTGSLNG TTMTLPATTL

         70          80          90         100         110         120
          |           |           |           |           |           |
TLSGHYATIS LLTVSGAWAK QMFTCRVAHT PSSTDWVDNK TFSVCSRDFT PPTVKILQSS

        130         140         150         160         170         180
          |           |           |           |           |           |
CDGGGHFPPT IQLLCLVSGY TPGTINITWL EDGQVMDVDL STASTTQEGE LASTQSELTL

        190         200         210         220         230         240
          |           |           |           |           |           |
SQKHWLSDRT YTCQVTYQGH TFEDSTKKCA DSNPRGVSAY LSRPSPFDLF IRKSPTITCL

        250         260         270         280         290         300
          |           |           |           |           |           |
VVDLAPSKGT VNLTWSRASG KPVNHSTRKE EKQRNGTLTV TSTLPVGTRD WIEGETYQCR

        310         320         330         340         350         360
          |           |           |           |           |           |
VTHPHLPRAL MRSTTKTSGP RAAPEVYAFA TPEWPGSRDK RTLACLIQNF MPEDISVQWL

        370         380         390         400         410         420
          |           |           |           |           |           |
HNEVQLPDAR HSTTQPRKTK GSGFFVFSRL EVTRAEWEQK DEFICRAVHE AASPSQTVQR

    428
     |
```

FIGURE 7 FF

AVSVNPGK

CODING SEQUENCE

```
atggactgga cctggatcct cttcttggtg gcagcagcca cgcgagtcca ctcccagacg      60
cagttggtgc agtctggggc tgaggtgagg aagcctgggg catcagtgag ggtctcctgc     120
aaggcttctg gatacacctt catcgactcc tatatccact ggatacgaca ggcccctggg     180
cacgggcttg agtgggtggg atggatcaac cctaacagtg gtggcacaaa ctatgctccg     240
agatttcagg gcagggtcac catgaccaga gacgcgtcct tcagtacagc ctacatggac     300
ctgagaagtc tgagatctga cgactcggcc gtgttttact gtgcgaaaag tgacctttt      360
tggagtgatt attataactt tgactactcg tacactttgg acgtctgggg ccaagggacc     420
acggtcaccg tctcctcagc ctccacacag agcccatccg tcttccccct gaccgctgc      480
tgcaaaaaca ttccctccaa tgccacctcc gtgactctgg gctgcctggc cacgggctac     540
ttcccggagc cggtgatggt gacctgggac acaggctccc tcaacgggac aactatgacc     600
ttaccagcca ccaccctcac gctctctggt cactatgcca ccatcagctt gctgaccgtc     660
tcgggtgcgt gggccaagca gatgttcacc tgccgtgtgg cacacactcc atcgtccaca     720
gactgggtcg acaacaaaac cttcagcgtc tgctccaggg acttcacccc gcccaccgtg     780
aagatcttac agtcgtcctg cgacggcggc gggcacttcc ccccgaccat ccagctcctg     840
tgcctcgtct ctgggtacac cccagggact atcaacatca cctggctgga ggacgggcag     900
gtcatggacg tggacttgtc caccgcctct accacgcagg agggtgagct ggcctccaca     960
caaagcgagc tcaccctcag ccagaagcac tggctgtcag accgcaccta cacctgccag    1020
gtcacctatc aaggtcacac ctttgaggac agcaccaaga agtgtgcaga ttccaacccg    1080
agaggggtga gcgcctacct aagccggccc agcccgttcg acctgttcat ccgcaagtcg    1140
cccacgatca cctgtctggt ggtggacctg gcacccagca aggggaccgt gaacctgacc    1200
tggtcccggg ccagtgggaa gcctgtgaac cactccacca gaaaggagga gaagcagcgc    1260
aatggcacgt taaccgtcac gtccacccctg ccggtgggca cccgagactg gatcgagggg    1320
gagacctacc agtgcagggt gacccacccc cacctgccca gggccctcat gcggtccacg    1380
accaagacca gcggcccgcg tgctgccccg gaagtctatg cgtttgcgac gccggagtgg    1440
ccggggagcc gggacaagcg caccctcgcc tgcctgatcc agaacttcat gcctgaggac    1500
atctcggtgc agtggctgca caacgaggtg cagctcccgg acgcccggca cagcacgacg    1560
cagccccgca agaccaaggg ctccggcttc ttcgtcttca gcgcctgga ggtgaccagg    1620
gccgaatggg agcagaaaga tgagttcatc tgccgtgcag tccatgaggc agcgagcccc    1680
tcacagaccg tccagcgagc ggtgtctgta aatcccggta aatga                    1725
```

GenBank

L00022. Human Ig active he...

```
LOCUS       HUMIGHAE2    1920 bp    DNA             PRI       22-DEC-1994
DEFINITION  Human Ig active heavy chain epsilon-1 gene, constant region.
ACCESSION   L00022 J00227 V00555
VERSION     L00022.1  GI:185035
KEYWORDS    C-region; epsilon-immunoglobulin; immunoglobulin heavy chain;
            processed gene.
SOURCE      Human myeloma cell line 266B1 DNA and cDNA to mRNA, clones
            H-Ig-epsilon-11, lambda-epsilon-1.2, pJJ71, pGET2 and K85/A12 (see
            comment).
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1920)
  AUTHORS   Flanagan,J.G. and Rabbitts,T.H.
  TITLE     The sequence of a human immunoglobulin epsilon heavy chain constant
            region gene, and evidence for three non-allelic genes
  JOURNAL   EMBO J. 1 (5), 655-660 (1982)
  MEDLINE   84236029
REFERENCE   2  (bases 528 to 736; 1044 to 1138)
  AUTHORS   Nishida,Y., Miki,T., Hisajima,H. and Honjo,T.
  TITLE     Cloning of human immunoglobulin epsilon chain genes: evidence for
            multiple C epsilon genes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 79 (12), 3833-3837 (1982)
  MEDLINE   82247945
REFERENCE   3  (bases 1 to 1920)
  AUTHORS   Kenten,J.H., Nolgaard,H.V., Houghton,M., Derbyshire,R.B., Viney,J.,
            Bell,L.O. and Gould,H.J.
  TITLE     Cloning and sequence determination of the gene for the human
```

FIGURE 7 GG

```
            immunoglobulin epsilon chain expressed in a myeloma cell line
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 79 (21), 6661-6665 (1982)
  MEDLINE   83065234
REFERENCE   4  (bases 98 to 1884)
  AUTHORS   Seno,M., Kurokawa,T., Ono,Y., Onda,H., Sasada,R., Igarashi,K.,
            Kikuchi,M., Sugino,Y., Nishida,Y. and Honjo,T.
  TITLE     Molecular cloning and nucleotide sequencing of human immunoglobulin
            epsilon chain cDNA
  JOURNAL   Nucleic Acids Res. 11 (3), 719-726 (1983)
  MEDLINE   83168897
REFERENCE   5  (bases 691 to 807; 1571 to 1818; 1860 to 1885)
  AUTHORS   Liu,F.T., Albrandt,K.A., Bry,C.G. and Ishizaka,T.
  TITLE     Expression of a biologically active fragment of human IgE epsilon
            chain in Escherichia coli
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 81 (17), 5369-5373 (1984)
  MEDLINE   84298140
COMMENT     [2] and [1] report the isolation of two other epsilon genes,
            epsilon-2 and epsilon-3. The authors in [2] claim that epsilon-3 is
            a pseudogene. Compared in [4] with the germline C-region sequence
            by Max, et al (Cell 29, 691-699 (1982)), and there are three
            nucleotide differences. The deduced amino acid sequence in [4]
            differs somewhat from the published C-region sequence. [5] produced
            expression of IgE in E.coli by insertion into expression vector
            pUC7.
            Complete source information:
            Human myeloma cell line 266B1 DNA [2],[1],[5] and cDNA to mRNA [3],
            [4], clones H-Ig-epsilon-11 [2], lambda-epsilon-1.2 [1], pJJ71 [3],
            pGET2 [4] and K85/A12 [5].
FEATURES                Location/Qualifiers
     source             1..1920
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /map="14q32.33"
     prim_transcript    <1..1886
                        /note="epsilon-1 mRNA"
     intron             <1..97
                        /gene="IGHE"
                        /note="epsilon-1 intron J-C"
     exon               98..406
                        /gene="IGHE"
                        /note="Ig heavy chain epsilon-1 (CH1 domain); G00-119-335"
     intron             407..613
                        /gene="IGHE"
                        /note="epsilon-1 intron A"
     exon               614..934
                        /gene="IGHE"
                        /note="Ig heavy chain epsilon-1 (CH2 domain)"
     conflict           735
                        /gene="IGHE"
                        /citation=[3]
                        /replace=""
     intron             935..1020
                        /gene="IGHE"
                        /note="epsilon-1 intron B"
     exon               1021..1344
                        /gene="IGHE"
                        /note="Ig heavy chain epsilon-1 (CH3 domain)"
     conflict           1124
                        /gene="IGHE"
                        /citation=[3]
                        /replace=""
     conflict           1337
                        /gene="IGHE"
```

FIGURE 7 HH

```
                     /citation=[3]
                     /replace=""
     intron          1345..1427
                     /gene="IGHE"
                     /note="epsilon-1 intron C"
     exon            1428..>1759
                     /note="Ig heavy chain epsilon-1 (CH4 domain)"
     conflict        1444..1445
                     /gene="IGHE"
                     /citation=[3]
                     /replace=""
     conflict        1612
                     /gene="IGHE"
                     /citation=[3]
                     /replace=""
     conflict        1785
                     /citation=[3]
                     /replace=""
     gene            join(L00021.1:57..495,1..1758)
                     /gene="IGHE"
     CDS             join(L00021.1:57..495,98..406,614..934,1021..1344,
                     1428..1759)
                     /partial
                     /gene="IGHE"
                     /note="Ig heavy chain epsilon-1 (V-D-J region)"
                     /codon_start=1
                     /protein_id="AAB59424.1"
                     /db_xref="GI:386807"
                     /db_xref="GDB:G00-119-335"
                     /translation="MDWTWILFLVAAATRVHSQTQLVQSGAEVRKPGASVRVSCKASG
                     YTFIDSYIHWIRQAPGHGLEWVGWINPNSGGTNYAPRFQGRVTMTRDASFSTAYMDLR
                     SLRSDDSAVFYCAKSDPFWSDYYNFDYSYTLDVWGQGTTVTVSSASTQSPSVFPLTRC
                     CKNIPSNATSVTLGCLATGYFPEPVMVTWDTGSLNGTTMTLPATTLTLSGHYATISLL
                     TVSGAWAKQMFTCRVAHTPSSTDWVDNKTFSVCSRDFTPPTVKILQSSCDGGGHFPPT
                     IQLLCLVSGYTPGTINITWLEDGQVMDVDLSTASTTQEGELASTQSELTLSQKHWLSD
                     RTYTCQVTYQGHTFEDSTKKCADSNPRGVSAYLSRPSPFDLFIRKSPTITCLVVDLAP
                     SKGTVNLTWSRASGKPVNHSTRKEEKQRNGTLTVTSTLPVGTRDWIEGETYQCRVTHP
                     HLPRALMRSTTKTSGPRAAPEVYAFATPEWPGSRDKRTLACLIQNFMPEDISVQWLHN
                     EVQLPDARHSTTQPRKTKGSGFFVFSRLEVTRAEWEQKDEFICRAVHEAASPSQTVQR
                     AVSVNPGK"
BASE COUNT      387 a    658 c    576 g    299 t
ORIGIN      About 3 kb after ; 1 bp upstream of BamHI site.
        1 ggatccctgc cacggggtcc ccagctcccc catccaggcc ccccaggctg atgggcgctg
       61 gcctgaggct ggcactgact aggttctgtc ctcacagcct ccacacagag cccatccgtc
      121 ttccccttga cccgctgctg caaaaacatt ccctccaatg ccacctccgt gactctgggc
      181 tgcctggcca cgggctactt cccggagccg gtgatggtga cctgggacac aggctccctc
      241 aacgggacaa ctatgacctt accagccacc accctcacgc tctctggtca ctatgccacc
      301 atcagcttgc tgaccgtctc gggtgcgtgg gccaagcaga tgttcacctg ccgtgtggca
      361 cacactccat cgtccacaga ctgggtcgac aacaaaacct tcagcggtaa gagagggcca
      421 agctcagaga ccacagttcc caggagtgcc aggctgaggg ctggcagagt gggcaggggt
      481 tgagggggtg ggtgggctca aacgtgggaa cacccagcat gcctggggac ccgggccagg
      541 acgtgggggc aagaggaggg cacacagagc tcagagaggc caacaaccct catgaccacc
      601 agctctcccc cagtctgctc cagggacttc accccgccca ccgtgaagat cttacagtcg
      661 tcctgcgacg gcggcgggca cttccccccg accatccagc tcctgtgcct cgtctctggg
      721 tacaccccag ggactatcaa catcacctgg ctggaggacg ggcaggtcat ggacgtggac
      781 ttgtccaccg cctctaccac gcaggagggt gagctggcct ccacacaaag cgagctcacc
      841 ctcagccaga agcactggct gtcagaccgc acctacacct gccaggtcac ctatcaaggt
      901 cacacctttg aggacagcac caagaagtgt gcaggtacgt tcccacctgc cctggtggcc
      961 gccacggagg ccagagaaga ggggcgggtg ggcctcacac agccctccgg tgtaccacag
     1021 attccaaccc gagaggggtg agcgcctacc taagccggcc cagcccgttc gacctgttca
     1081 tccgcaagtc gcccacgatc acctgtctgg tggtggacct ggcacccagc aaggggaccg
     1141 tgaacctgac ctggtcccgg gccagtggga agcctgtgaa ccactccacc agaaaggagg
     1201 agaagcagcg caatggcacg ttaaccgtca cgtccaccct gccggtgggc acccgagact
```

FIGURE 7 II

```
1261 ggatcgaggg ggagacctac cagtgcaggg tgacccaccc ccacctgccc agggccctca
1321 tgcggtccac gaccaagacc agcggtgagc catgggcagg ccggggtcgt gggggaaggg
1381 agggagcgag tgagcggggc ccgggctgac cccacgtctg gccacaggcc cgcgtgctgc
1441 cccggaagtc tatgcgtttg cgacgccgga gtggccgggg agccgggaca agcgcaccct
1501 cgcctgcctg atccagaact tcatgcctga ggacatctcg gtgcagtggc tgcacaacga
1561 ggtgcagctc ccggacgccc ggcacagcac gacgcagccc cgcaagacca agggctccgg
1621 cttcttcgtc ttcagccgcc tggaggtgac cagggccgaa tgggagcaga aagatgagtt
1681 catctgccgt gcagtccatg aggcagcgag ccccttcacag accgtccagc gagcggtgtc
1741 tgtaaatccc ggtaaatgac gtactcctgc ctccctccct cccagggctc catccagctg
1801 tgcagtgggg aggactggcc agaccttctg tccactgttg caatgacccc aggaagctac
1861 ccccaataaa ctgtgcctgc tcagagcccc agtacaccca ttcttgggag cgggcagggc
```

IX. HUMAN IG MU CHAIN C REGION - HOMO SAPIENS (HUMAN).

AMINO ACID SEQUENCE

>sp|P01871|MUC_HUMAN IG MU CHAIN C REGION - Homo sapiens (Human).

```
        10         20         30         40         50         60
         |          |          |          |          |          |
GSASAPTLFP LVSCENSPSD TSSVAVGCLA QDFLPDSITF SWKYKNNSDI SSTRGFPSVL

        70         80         90        100        110        120
         |          |          |          |          |          |
RGGKYAATSQ VLLPSKDVMQ GTDEHVVCKV QHPNGNKEKN VPLPVIAELP PKVSVFVPPR

       130        140        150        160        170        180
         |          |          |          |          |          |
DGFFGNPRSK SKLICQATGF SPRQIQVSWL REGKQVGSGV TTDQVQAEAK ESGPTTYKVT

       190        200        210        220        230        240
         |          |          |          |          |          |
STLTIKESDW LSQSMFTCRV DHRGLTFQQN ASSMCVPDQD TAIRVFAIPP SFASIFLTKS

       250        260        270        280        290        300
         |          |          |          |          |          |
TKLTCLVTDL TTYDSVTISW TRQNGEAVKT HTNISESHPN ATFSAVGEAS ICEDDWNSGE

       310        320        330        340        350        360
         |          |          |          |          |          |
RFTCTVTHTD LPSPLKQTIS RPKGVALHRP DVYLLPPARE QLNLRESATI TCLVTGFSPA

       370        380        390        400        410        420
         |          |          |          |          |          |
DVFVQWMQRG QPLSPEKYVT SAPMPEPQAP GRYFAHSILT VSEEEWNTGE TYTCVVAHEA

       430        440        450
         |          |          |
LPNRVTERTV DKSTGKPTLY NVSLVMSDTA GTCY
```

CODING SEQUENCE

```
atggactgga cctggaggtt cctctttgtg gtggcagcag ctacaggtgt ccagtcccag      60
gtgcagctgg tgcagtctgg ggctgaggtg aagaagcctg ggtcctcggt gaaggtctcc     120
tgcaaggctt ctggaggcac cttcagcagc tatgctatca gctgggtgcg acaggcccct     180
ggacaagggc ttgagtggat gggagggatc atccctatct ttggtacagc aaactacgca     240
cagaagttcc agggcagagt cacgattacc gcggacgaat ccacgagcac agcctacatg     300
gagctgagca gcctgagatc tgaggacacg gccgtgtatt actgtgcgaa aaccgggatc     360
ctggggccgt atagcagtgg ctggtacccg aactcggact actactacta cggtatggac     420
gtctgggggc aagggaccac ggtcaccgtc tcctcaggga gtgcatccgc cccaaccctt     480
ttccccctcg tctcctgtga gaattccccg tcggatacga gcagcgtggc cgttggctgc     540
ctcgcacagg acttccttcc cgactccatc actttctcct ggaaatacaa gaacaactct     600
gacatcagca gcacccgggg cttcccatca gtcctgagag ggggcaagta cgcagccacc     660
```

FIGURE 7 JJ

```
tcacaggtgc tgctgccttc caaggacgtc atgcagggca cagacgaaca cgtggtgtgc      720
aaagtccagc accccaacgg caacaaagaa aagaacgtgc ctcttccagt gattgctgag      780
ctgcctccca aagtgagcgt cttcgtccca ccccgcgacg gcttcttcgg caacccccgc      840
agcaagtcca agctcatctg ccaggccacg ggtttcagtc cccggcagat tcaggtgtcc      900
tggctgcgcg aggggaagca ggtggggtct ggcgtcacca cggaccaggt gcaggctgag      960
gccaaagagt ctgggcccac gacctacaag gtgaccagca cactgaccat caaagagagc     1020
gactggctca gccagagcat gttcacctgc cgcgtggatc acaggggcct gaccttccag     1080
cagaatgcgt cctccatgtg tgtccccgat caagacacag ccatccgggt cttcgccatc     1140
cccccatcct ttgccagcat cttcctcacc aagtccacca agttgacctg cctggtcaca     1200
gacctgacca cctatgacag cgtgaccatc tcctggaccc gccagaatgg cgaagctgtg     1260
aaaacccaca ccaacatctc cgagagccac cccaatgcca ctttcagcgc cgtgggtgag     1320
gccagcatct gcgaggatga ctggaattcc ggggagaggt tcacgtgcac cgtgacccac     1380
acagacctgc cctcgccact gaagcagacc atctcccggc ccaaggggct ggccctgcac     1440
aggcccgatg tctacttgct gccaccagcc cgggagcagc tgaacctgcg ggagtcggcc     1500
accatcacgt gcctggtgac gggcttctct cccgcggacg tcttcgtgca gtggatgcag     1560
aggggcagc ccttgtcccc ggagaagtat gtgaccagcg ccccaatgcc tgagccccag     1620
gccccaggcc ggtacttcgc ccacagcatc ctgaccgtgt ccgaagagga atggaacacg     1680
ggggagacct acacctgcgt ggtggcccat gaggccctgc ccaacagggt caccgagagg     1740
accgtggaca agtccaccga gggggaggtg agcgccgacg aggagggctt tgagaacctg     1800
tgggccaccg cctccacctt catcgtcctc ttcctcctga gcctcttcta cagtaccacc     1860
gtcaccttgt tcaaggtgaa atga                                            1884
```

GenBank

X17115. Human mRNA for IgM

LOCUS	HSIGM201 2213 bp mRNA PRI 03-APR-1995
DEFINITION	Human mRNA for IgM heavy chain complete sequence.
ACCESSION	X17115
VERSION	X17115.1 GI:33450
KEYWORDS	Ig heavy chain; IgM gene; IgM heavy chain; transmembrane protein.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 2213)
AUTHORS	Friedlander,R.M.
TITLE	Direct Submission
JOURNAL	Submitted (03-NOV-1989) Friedlander R. M., Harvard Medical School, Howard Hughes Medical Institute, Department of Genetics, 25 Shattuck St, Boston, MA 02115, USA
REFERENCE	2 (bases 1 to 2213)
AUTHORS	Friedlander,R.M., Nussenzweig,M.C. and Leder,P.
TITLE	Complete nucleotide sequence of the membrane form of the human IgM heavy chain
JOURNAL	Nucleic Acids Res. 18 (14), 4278 (1990)
MEDLINE	90332450
REFERENCE	3 (bases 1 to 2213)
AUTHORS	Kristensen,T., Lopez,R. and Prydz,H.
TITLE	An estimate of the sequencing error frequency in the DNA sequence databases
JOURNAL	DNA Seq. 2 (6), 343-346 (1992)
MEDLINE	93075997
REMARK	Erratum:[[published erratum appears in DNA Seq 1993;3(5):337]]
COMMENT	For genomic sequence see <K01306>,<X14939> and <X14940>. The author reports various conflicts with these sequences. Data kindly reviewed (30-MAY-1990) by Friedlander R.M.

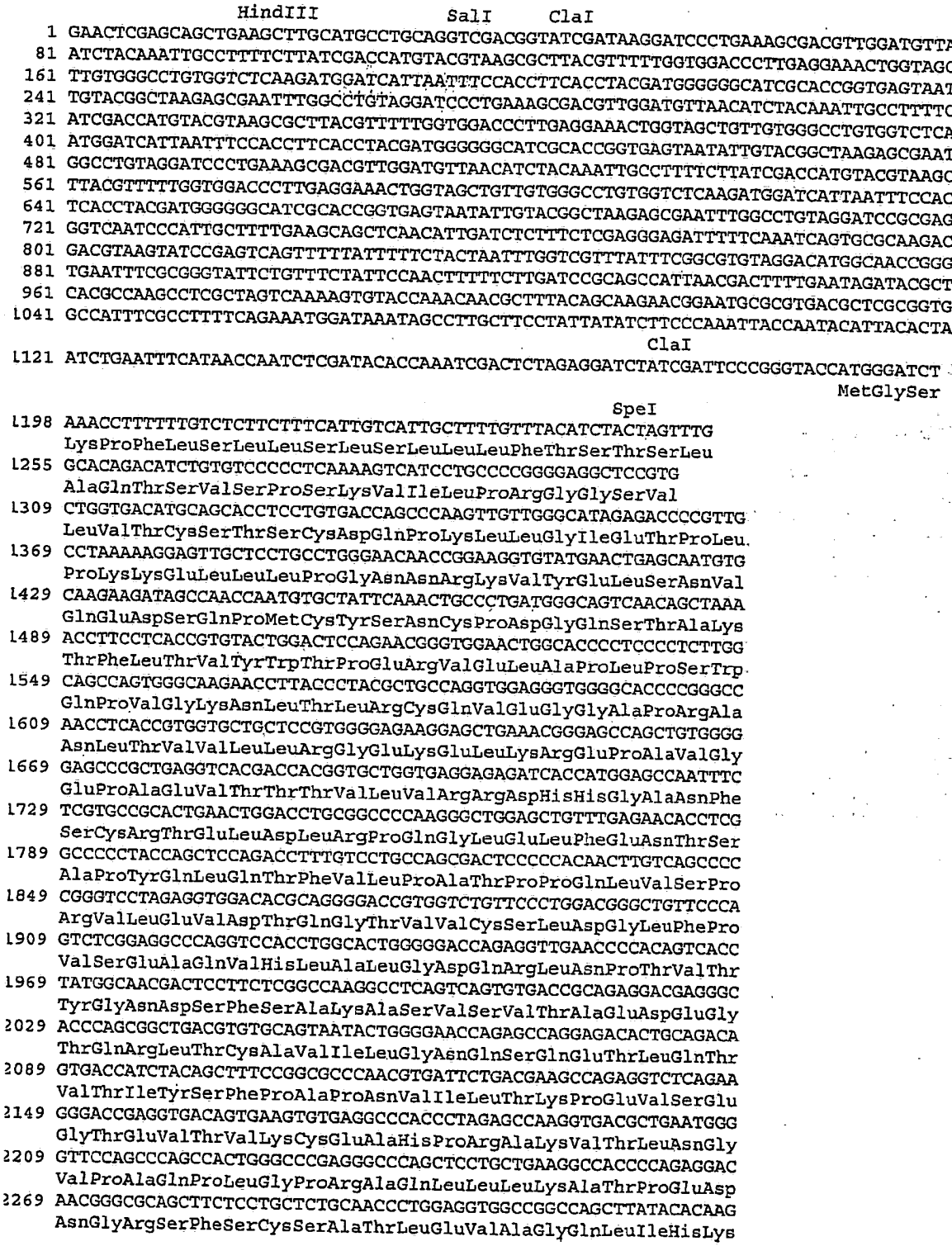
FEATURES	Location/Qualifiers
source	1..2213
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="201-203"
	/cell_line="lymphoma 201"
	/cell_type="B"

FIGURE 7 KK

```
                     /tissue_type="lymphoid"
     misc_feature    1..39
                     /note="putative VECTOR sequence Bluescript SKP+"
                     /citation=[3]
     CDS             73..1956
                     /note="precursor (AA -15 to 612)"
                     /codon_start=1
                     /protein_id="CAA34971.1"
                     /db_xref="GI:33451"
                     /db_xref="SWISS-PROT:P01871"
                     /db_xref="SWISS-PROT:P20769"
                     /translation="MDWTWRFLFVVAAATGVQSQVQLVQSGAEVKKPGSSVKVSCKAS
                     GGTFSSYAISWVRQAPGQGLEWMGGIIPIFGTANYAQKFQGRVTITADESTSTAYMEL
                     SSLRSEDTAVYYCAKTGILGPYSSGWYPNSDYYYYGMDVWGQGTTVTVSSGSASAPTL
                     FPLVSCENSPSDTSSVAVGCLAQDFLPDSITFSWKYKNNSDISSTRGFPSVLRGGKYA
                     ATSQVLLPSKDVMQGTDEHVVCKVQHPNGNKEKNVPLPVIAELPPKVSVFVPPRDGFF
                     GNPRSKSKLICQATGFSPRQIQVSWLREGKQVGSGVTTDQVQAEAKESGPTTYKVTST
                     LTIKESDWLSQSMFTCRVDHRGLTFQQNASSMCVPDQDTAIRVFAIPPSFASIFLTKS
                     TKLTCLVTDLTTYDSVTISWTRQNGEAVKTHTNISESHPNATFSAVGEASICEDDWNS
                     GERFTCTVTHTDLPSPLKQTISRPKGVALHRPDVYLLPPAREQLNLRESATITCLVTG
                     FSPADVFVQWMQRGQPLSPEKYVTSAPMPEPQAPGRYFAHSILTVSEEEWNTGETYTC
                     VVAHEALPNRVTERTVDKSTEGEVSADEEGFENLWATASTFIVLFLLSLFYSTTVTLF
                     KVK"
     sig_peptide     73..117
     mat_peptide     118..1953
                     /product="IgM heavy chain (AA 1 to 612)"
     polyA_site      2213
                     /note="polyA site"
BASE COUNT      462 a    708 c    629 g    414 t
ORIGIN
        1 gctctagaac tagtggatcc cccgggctgc aggaattctc taaagaagcc cctgggagca
       61 cagctcatca ccatggactg gacctggagg ttcctctttg tggtggcagc agctacaggt
      121 gtccagtccc aggtgcagct ggtgcagtct ggggctgagg tgaagaagcc tgggtcctcg
      181 gtgaaggtct cctgcaaggc ttctggaggc accttcagca gctatgctat cagctgggtg
      241 cgacaggccc ctggacaagg gcttgagtgg atgggaggga tcatccctat ctttggtaca
      301 gcaaactacg cacagaagtt ccagggcaga gtcacgatta ccgcggacga atccacgagc
      361 acagcctaca tggagctgag cagcctgaga tctgaggaca cggccgtgta ttactgtgcg
      421 aaaaccggga tcctggggcc gtatagcagt ggctggtacc cgaactcgga ctactactac
      481 tacggtatgg acgtctgggg ccaagggacc acggtcaccg tctcctcagg gagtgcatcc
      541 gccccaaccc ttttcccccc cgtctcctgt gagaattccc cgtcggatac gagcagcgtg
      601 gccgttggct gcctcgcaca ggacttcctt cccgactcca tcactttctc ctggaaatac
      661 aagaacaact ctgacatcag cagcacccgg ggcttcccat cagtcctgag agggggcaag
      721 tacgcagcca cctcacaggt gctgctgcct tccaaggacg tcatgcaggg cacagacgaa
      781 cacgtggtgt gcaaagtcca gcaccccaac ggcaacaaag aaaagaacgt gcctcttcca
      841 gtgattgctg agctgcctcc caaagtgagc gtcttcgtcc caccccgcga cggcttcttc
      901 ggcaaccccc gcagcaagtc caagctcatc tgccaggcca cgggtttcag tccccggcag
      961 attcaggtgt cctggctgcg cgaggggaag caggtggggt ctggcgtcac cacggaccag
     1021 gtgcaggctg aggccaaaga gtctgggccc acgacctaca aggtgaccag cacactgacc
     1081 atcaaagaga gcgactggct cagccagagc atgttcacct gccgcgtgga tcacaggggc
     1141 ctgaccttcc agcagaatgc gtcctccatg tgtgtccccg atcaagacac agccatccgg
     1201 gtcttcgcca tccccccatc ctttgccagc atcttcctca ccaagtccac caagttgacc
     1261 tgcctggtca cagacctgac cacctatgac agcgtgacca tctcctggac ccgccagaat
     1321 ggcgaagctg tgaaaaccca caccaacatc tccgagagcc accccaatgc cactttcagc
     1381 gccgtgggtg aggccagcat ctgcgaggat gactggaatt ccggggagag gttcacgtgc
     1441 accgtgaccc acacagacct gccctcgcca ctgaagcaga ccatctcccg gcccaagggg
     1501 gtggccctgc acaggcccga tgtctacttg ctgccaccag cccgggagca gctgaacctg
     1561 cgggagtcgg ccaccatcac gtgcctggtg acgggcttct ctcccgcgga cgtcttcgtg
     1621 cagtggatgc agaggggcga gcccttgtcc ccggagaagt atgtgaccag cgccccaatg
     1681 cctgagcccc aggccccagg ccggtacttc gcccacagca tcctgaccgt gtccgaagag
     1741 gaatggaaca cgggggagac ctacacctgc gtggtggccc atgaggccct gcccaacagg
     1801 gtcaccgaga ggaccgtgga caagtccacc gaggggggagg tgagcgccga cgaggagggc
     1861 tttgagaacc tgtgggccac cgcctccacc ttcatcgtcc tcttcctcct gagcctcttc
     1921 tacagtacca ccgtcacctt gttcaaggtg aaatgatccc aacagaagaa catcggagac
```

FIGURE 7 LL

```
1981 cagagagagg aactcaaagg ggcgctgcct ccgggtctgg ggtcctggcc tgcgtggcct
2041 gttggcacgt gtttctcttc ccgcccggcc tccagttgtg tgctctcaca caggcttcct
2101 tctcgaccgg caggggctgg ctggcttgca ggccacgagg tgggctctac cccacactgc
2161 tttgctgtgt atacgcttgt tgccctgaaa taaatatgca cattttatcc atg
```




```
                  HindIII           SalI      ClaI
   1 GAACTCGAGCAGCTGAAGCTTGCATGCCTGCAGGTCGACGGTATCGATAAGGATCCCTGAAAGCGACGTTGGATGTTAAC
  81 ATCTACAAATTGCCTTTTCTTATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGACCCTTGAGGAAACTGGTAGCTG
 161 TTGTGGGCCTGTGGTCTCAAGATGGATCATTAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGTAATAT
 241 TGTACGGCTAAGAGCGAATTTGGCCTGTAGGATCCCTGAAAGCGACGTTGGATGTTAACATCTACAAATTGCCTTTTCTT
 321 ATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGACCCTTGAGGAAACTGGTAGCTGTTGTGGGCCTGTGGTCTCAAG
 401 ATGGATCATTAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGTAATATTGTACGGCTAAGAGCGAATTT
 481 GGCCTGTAGGATCCCTGAAAGCGACGTTGGATGTTAACATCTACAAATTGCCTTTTCTTATCGACCATGTACGTAAGCGC
 561 TTACGTTTTTGGTGGACCCTTGAGGAAACTGGTAGCTGTTGTGGGCCTGTGGTCTCAAGATGGATCATTAATTTCCACCT
 641 TCACCTACGATGGGGGGCATCGCACCGGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCCTGTAGGATCCGCGAGCT
 721 GGTCAATCCCATTGCTTTTGAAGCAGCTCAACATTGATCTCTTTCTCGAGGGAGATTTTTCAAATCAGTGCGCAAGACGT
 801 GACGTAAGTATCCGAGTCAGTTTTTATTTTTCTACTAATTTGGTCGTTTATTTCGGCGTGTAGGACATGGCAACCGGGCC
 881 TGAATTTCGCGGGTATTCTGTTTCTATTCCAACTTTTTCTTGATCCGCAGCCATTAACGACTTTTGAAATAGATACGCTGA
 961 CACGCCAAGCCTCGCTAGTCAAAAGTGTACCAAACAACGCTTTACAGCAAGAACGGAATGCGCGTGACGCTCGCGGTGAC
1041 GCCATTTCGCCTTTTCAGAAATGGATAAATAGCCTTGCTTCCTATTATATCTTCCCAAATTACCAATACATTACACTAGC
                                                    ClaI
1121 ATCTGAATTTCATAACCAATCTCGATACACCAAATCGACTCTAGAGGATCTATCGATTCCCGGGTACCATGGGATCT
                                                                        MetGlySer
                                                  SpeI
1198 AAACCTTTTTTGTCTCTTCTTTCATTGTCATTGCTTTTGTTTACATCTACTAGTTTG
     LysProPheLeuSerLeuLeuSerLeuSerLeuLeuLeuPheThrSerThrSerLeu
1255 GCACAGACATCTGTGTCCCCCTCAAAAGTCATCCTGCCCCGGGGAGGCTCCGTG
     AlaGlnThrSerValSerProSerLysValIleLeuProArgGlySerVal
1309 CTGGTGACATGCAGCACCTCCTGTGACCAGCCCAAGTTGTTGGGCATAGAGACCCCGTTG
     LeuValThrCysSerThrSerCysAspGlnProLysLeuLeuGlyIleGluThrProLeu
1369 CCTAAAAAGGAGTTGCTCCTGCCTGGGAACAACCGGAAGGTGTATGAACTGAGCAATGTG
     ProLysLysGluLeuLeuProGlyAsnAsnArgLysValTyrGluLeuSerAsnVal
1429 CAAGAAGATAGCCAACCAATGTGCTATTCAAACTGCCCTGATGGGCAGTCAACAGCTAAA
     GlnGluAspSerGlnProMetCysTyrSerAsnCysProAspGlyGlnSerThrAlaLys
1489 ACCTTCCTCACCGTGTACTGGACTCCAGAACGGGTGGAACTGGCACCCCTCCCCTCTTGG
     ThrPheLeuThrValTyrTrpThrProGluArgValGluLeuAlaProLeuProSerTrp
1549 CAGCCAGTGGGCAAGAACCTTACCCTACGCTGCCAGGTGGAGGGTGGGGCACCCCGGGCC
     GlnProValGlyLysAsnLeuThrLeuArgCysGlnValGluGlyGlyAlaProArgAla
1609 AACCTCACCGTGGTGCTGCTCCGTGGGGAGAAGGAGCTGAAACGGGAGCCAGCTGTGGGG
     AsnLeuThrValValLeuLeuArgGlyGluLysGluLeuLysArgGluProAlaValGly
1669 GAGCCCGCTGAGGTCACGACCACGGTGCTGGTGAGGAGAGATCACCATGGAGCCAATTTC
     GluProAlaGluValThrThrThrValLeuValArgArgAspHisHisGlyAlaAsnPhe
1729 TCGTGCCGCACTGAACTGGACCTGCGGCCCCAAGGGCTGGAGCTGTTTGAGAACACCTCG
     SerCysArgThrGluLeuAspLeuArgProGlnGlyLeuGluLeuPheGluAsnThrSer
1789 GCCCCCTACCAGCTCCAGACCTTTGTCCTGCCAGCGACTCCCCCACAACTTGTCAGCCCC
     AlaProTyrGlnLeuGlnThrPheValLeuProAlaThrProProGlnLeuValSerPro
1849 CGGGTCCTAGAGGTGGACACGCAGGGGACCGTGGTCTGTTCCCTGGACGGGCTGTTCCCA
     ArgValLeuGluValAspThrGlnGlyThrValValCysSerLeuAspGlyLeuPhePro
1909 GTCTCGGAGGCCCAGGTCCACCTGGCACTGGGGGACCAGAGGTTGAACCCCACAGTCACC
     ValSerGluAlaGlnValHisLeuAlaLeuGlyAspGlnArgLeuAsnProThrValThr
1969 TATGGCAACGACTCCTTCTCGGCCAAGGCCTCAGTCAGTGTGACCGCAGAGGACGAGGGC
     TyrGlyAsnAspSerPheSerAlaLysAlaSerValSerValThrAlaGluAspGluGly
2029 ACCCAGCGGCTGACGTGTGCAGTAATACTGGGGAACCAGAGCCAGGAGACACTGCAGACA
     ThrGlnArgLeuThrCysAlaValIleLeuGlyAsnGlnSerGlnGluThrLeuGlnThr
2089 GTGACCATCTACAGCTTTCCGGCGCCCAACGTGATTCTGACGAAGCCAGAGGTCTCAGAA
     ValThrIleTyrSerPheProAlaProAsnValIleLeuThrLysProGluValSerGlu
2149 GGGACCGAGGTGACAGTGAAGTGTGAGGCCCACCCTAGAGCCAAGGTGACGCTGAATGGG
     GlyThrGluValThrValLysCysGluAlaHisProArgAlaLysValThrLeuAsnGly
2209 GTTCCAGCCCAGCCACTGGGCCCGAGGGCCCAGCTCCTGCTGAAGGCCACCCCAGAGGAC
     ValProAlaGlnProLeuGlyProArgAlaGlnLeuLeuLeuLysAlaThrProGluAsp
2269 AACGGGCGCAGCTTCTCCTGCTCTGCAACCCTGGAGGTGGCCGGCCAGCTTATACACAAG
     AsnGlyArgSerPheSerCysSerAlaThrLeuGluValAlaGlyGlnLeuIleHisLys
```

FIGURE 8 A - 1

```
2329 AACCAGACCCGGGAGCTTCGTGTCCTGTATGGCCCCCGACTGGACGAGAGGGATTGTCCG
     AsnGlnThrArgGluLeuArgValLeuTyrGlyProArgLeuAspGluArgAspCysPro
2389 GGAAACTGGACGTGGCCAGAAAATTCCCAGCAGACTCCAATGTGCCAGGCTTGGGGGAAC
     GlyAsnTrpThrTrpProGluAsnSerGlnGlnThrProMetCysGlnAlaTrpGlyAsn
               SacI
2449 CCATTGCCCGAGCTCAAGTGTCTAAAGGATGGCACTTTCCCACTGCCCATCGGGGAATCA
     ProLeuProGluLeuLysCysLeuLysAspGlyThrPheProLeuProIleGlyGluSer
2509 GTGACTGTCACTCGAGATCTTGAGGGCACCTACCTCTGTCGGGCCAGGAGCACTCAAGGG
     ValThrValThrArgAspLeuGluGlyThrTyrLeuCysArgAlaArgSerThrGlnGly
                                    SpeI     SacI
2569 GAGGTCACCCGCGAGGTGACCGTGAATGTGACTAGTGGGAGCTCAGCATCCCCG
     GluValThrArgGluValThrValAsnValThrSerGlySerSerAlaSerPro
2623 ACCAGCCCCAAGGTCTTCCCGCTGAGCCTCGACAGCACCCCCCAAGATGGGAACGTGGTC
     ThrSerProLysValPheProLeuSerLeuAspSerThrProGlnAspGlyAsnValVal
2683 GTCGCATGCCTGGTCCAGGGCTTCTTCCCCCAGGAGCCACTCAGTGTGACCTGGAGCGAA
     ValAlaCysLeuValGlnGlyPhePheProGlnGluProLeuSerValThrTrpSerGlu
2743 AGCGGACAGAACGTGACCGCCAGAAACTTCCCACCTAGCCAGGATGCCTCCGGGGACCTG
     SerGlyGlnAsnValThrAlaArgAsnPheProProSerGlnAspAlaSerGlyAspLeu
2803 TACACCACGAGCAGCCAGCTGACCCTGCCGGCCACACAGTGCCCAGACGGCAAGTCCGTG
     TyrThrThrSerSerGlnLeuThrLeuProAlaThrGlnCysProAspGlyLysSerVal
2863 ACATGCCACGTGAAGCACTACACGAATTCCAGCCAGGATGTGACTGTGCCCTGCCGAGTT
     ThrCysHisValLysHisTyrThrAsnSerSerGlnAspValThrValProCysArgVal
2923 CCCCCACCTCCCCCATGCTGCCACCCCCGACTGTCGCTGCACCGACCGGCCCTCGAGGAC
     ProProProProProCysCysHisProArgLeuSerLeuHisArgProAlaLeuGluAsp
2983 CTGCTCTTAGGTTCAGAAGCGAACCTCACGTGCACACTGACCGGCCTGAGAGATGCCTCT
     LeuLeuLeuGlySerGluAlaAsnLeuThrCysThrLeuThrGlyLeuArgAspAlaSer
3043 GGTGCCACCTTCACCTGGACGCCCTCAAGTGGGAAGAGCGCTGTTCAAGGACCACCTGAG
     GlyAlaThrPheThrTrpThrProSerSerGlyLysSerAlaValGlnGlyProProGlu
3103 CGTGACCTCTGTGGCTGCTACAGCGTGTCCAGAGTACTTCCTGGCTGTGCCCAGCCATGG
     ArgAspLeuCysGlyCysTyrSerValSerArgValLeuProGlyCysAlaGlnProTrp
3163 AACCATGGGGAGACCTTCACCTGCACTGCTGCCCACCCCGAGTTGAAGACCCCACTAACC
     AsnHisGlyGluThrPheThrCysThrAlaAlaHisProGluLeuLysThrProLeuThr
3223 GCCAACATCACAAAATCCGGAAACACATTCCGGCCCGAGGTCCACCTGCTGCCGCCGCCG
     AlaAsnIleThrLysSerGlyAsnThrPheArgProGluValHisLeuLeuProProPro
3283 TCGGAGGAGCTGGCCCTGAACGAGCTGGTGACGCTGACGTGCCTGGCACGTGGCTTCAGC
     SerGluGluLeuAlaLeuAsnGluLeuValThrLeuThrCysLeuAlaArgGlyPheSer
3343 CCCAAGGATGTGCTGGTTCGCTGGCTGCAGGGGTCACAGGAGCTGCCCCGCGAGAAGTAC
     ProLysAspValLeuValArgTrpLeuGlnGlySerGlnGluLeuProArgGluLysTyr
3403 CTGACTTGGGCATCCCGGCAGGAGCCCAGCCAGGGCACCACCACCTATGCTGTGACCAGC
     LeuThrTrpAlaSerArgGlnGluProSerGlnGlyThrThrThrTyrAlaValThrSer
3463 ATACTGCGCGTGGCAGCCGAGGACTGGAAGAAGGGGGAGACCTTCTCCTGCATGGTGGGC
     IleLeuArgValAlaAlaGluAspTrpLysLysGlyGluThrPheSerCysMetValGly
3523 CACGAGGCCCTGCCGCTGGCCTTCACACAGAAGACCATCGACCGCTTGGCGGGTAAACCC
     HisGluAlaLeuProLeuAlaPheThrGlnLysThrIleAspArgLeuAlaGlyLysPro
3583 ACCCATATCAATGTGTCTGTTGTCATGGCGGAGGCGGACGGCACCTGCTACAGATCTGAA
     ThrHisIleAsnValSerValValMetAlaGluAlaAspGlyThrCysTyrArgSerGlu
               junEcoRI marker
3643 AAGGATGAACTTTAGAATTCGATATCAAGCTAATTCCCGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGAT
     LysAspGluLeu
3718 TGAATCCTGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAA
3798 TGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAAT
3878 ATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGGGGATCTGCCGGTCTCCCTATAG
3958 TGAGTCGTATTAATTTCGATAAGCCAGGTTAACCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTAT
4038 TGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAA
4118 AGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGG
4198 AACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAG
4278 TCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTC
```

FIGURE 8 A – 2

```
4358 CGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGG
4438 TATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT
4518 ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTA
4598 GCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTT
4678 GGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGG
4758 TAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTA
4838 CGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAG
4918 ATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTT
4998 AATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTA
5078 CGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCA
5158 GCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTG
5238 TTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGT
5318 CACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGC
5398 AAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGC
5478 AGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCT
5558 GAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTA
5638 AAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTA
5718 ACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAA
5798 ATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATT
5878 TATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATT
5958 TCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGC
6038 CCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTG
6118 TAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGC
                                    NdeI
6198 GGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGGACATATTGTCGTTAGAACGCGGCTACAATTAATACATAACCT
6278 TATGTATCATACACATACGATTTAGGTGACACTATA
```

FIGURE 8 A – 3

Bean Legumin Signal Peptide

MetSerLysProPheLeuSerLeuLeuSerLeuSerLeuLeuLeuPheThrSerThrCysLeuAla

FIGURE 8 B

Nucleotide and Amino Acid Sequence of Protein Coding Region of pSHuJ

```
1141
      AGGATCTATCGATTCCCGGGTACC ATG GAG AAC CAT TTG CTT TTC TGG GGA GTC CTG GCG
                               met glu asn his leu leu phe trp gly val leu ala
1201/13
GTT TTT ATT AAG GCT GTT CAT GTG AAA GCC CAA GAA GAT GAA AGG ATT GTT CTT GTT GAC
val phe ile lys ala val his val lys ala gln glu asp glu arg ile val leu val asp
1261/33
AAC AAA TGT AAG TGT GCC CGG ATT ACT TCC AGG ATC ATC CGT TCT TCC GAA GAT CCT AAT
asn lys cys lys cys ala arg ile thr ser arg ile ile arg ser ser glu asp pro asn
1321/53
GAG GAC ATT GTG GAG AGA AAC ATC CGA ATT ATT GTT CCT CTG AAC AAC AGG GAG AAT ATC
glu asp ile val glu arg asn ile arg ile ile val pro leu asn asn arg glu asn ile
1381/73
TCT GAT CCC ACC TCA CCA TTG AGA ACC AGA TTT GTG TAC CAT TTG TCT GAC CTC TGT AAA
ser asp pro thr ser pro leu arg thr arg phe val tyr his leu ser asp leu cys lys
1441/93
AAA TGT GAT CCT ACA GAA GTG GAG CTG GAT AAT CAG ATA GTT ACT GCT ACC CAG AGC AAT
lys cys asp pro thr glu val glu leu asp asn gln ile val thr ala thr gln ser asn
1501/113
ATC TGT GAT GAA GAC AGT GCT ACA GAG ACC TGC TAC ACT TAT GAC AGA AAC AAG TGC TAC
ile cys asp glu asp ser ala thr glu thr cys tyr thr tyr asp arg asn lys cys tyr
1561/133
ACA GCT GTG GTC CCA CTC GTA TAT GGT GGT GAG ACC AAA ATG GTG GAA ACA GCC TTA ACC
thr ala val val pro leu val tyr gly gly glu thr lys met val glu thr ala leu thr
1621/153
CCA GAT GCC TGC TAT CCT GAC TGA ATTC
pro asp ala cys tyr pro asp
```

FIGURE 8 C

Nucleotide and Amino Acid Sequence of Protein Coding Region of pSHuSC-

```
1137
                                        GTCGATTCCCGGGTACC ATG GTG CTC TTC GTG CTC ACC TGC
                                                          met val leu phe val leu thr cys
1178/9
CTG CTG GCG GTC TTC CCA GCC ATC TCC ACG AAG AGT CCC ATA TTT GGT CCC GAG GAG GTG
leu leu ala val phe pro ala ile ser thr lys ser pro ile phe gly pro glu glu val
1238/29
AAT AGT GTG GAA GGT AAC TCA GTG TCC ATC ACG TGC TAC TAC CCA CCC ACC TCT GTC AAC
asn ser val glu gly asn ser val ser ile thr cys tyr tyr pro pro thr ser val asn
1298/49
CGG CAC ACC CGG AAG TAC TGG TGC CGG CAG GGA GCT AGA GGT GGC TGC ATA ACC CTC ATC
arg his thr arg lys tyr trp cys arg gln gly ala arg gly gly cys ile thr leu ile
1358/69
TCC TCG GAG GGC TAC GTC TCC AGC AAA TAT GCA GGC AGG GCT AAC CTC ACC AAC TTC CCG
ser ser glu gly tyr val ser ser lys tyr ala gly arg ala asn leu thr asn phe pro
1418/89
GAG AAC GGC ACA TTT GTG GTG AAC ATT GCC CAG CTG AGC CAG GAT GAC TCC GGG CGC TAC
glu asn gly thr phe val val asn ile ala gln leu ser gln asp asp ser gly arg tyr
1478/109
AAG TGT GGC CTG GGC ATC AAT AGC CGA GGC CTG TCC TTT GAT GTC AGC CTG GAG GTC AGC
lys cys gly leu gly ile asn ser arg gly leu ser phe asp val ser leu glu val ser
1538/129
CAG GGT CCT GGG CTC CTA AAT GAC ACT AAA GTC TAC ACA GTG GAC CTG GGC AGA ACG GTG
gln gly pro gly leu leu asn asp thr lys val tyr thr val asp leu gly arg thr val
1598/149
ACC ATC AAC TGC CCT TTC AAG ACT GAG AAT GCT CAA AAG AGG AAG TCC TTG TAC AAG CAG
thr ile asn cys pro phe lys thr glu asn ala gln lys arg lys ser leu tyr lys gln
1658/169
ATA GGC CTG TAC CCT GTG CTG GTC ATC GAC TCC AGT GGT TAT GTG AAT CCC AAC TAT ACA
ile gly leu tyr pro val leu val ile asp ser ser gly tyr val asn pro asn tyr thr
1718/189
GGA AGA ATA CGC CTT GAT ATT CAG GGT ACT GGC CAG TTA CTG TTC AGC GTT GTC ATC AAC
gly arg ile arg leu asp ile gln gly thr gly gln leu leu phe ser val val ile asn
1778/209
CAA CTC AGG CTC AGC GAT GCT GGG CAG TAT CTC TGC CAG GCT GGG GAT GAT TCC AAT AGT
gln leu arg leu ser asp ala gly gln tyr leu cys gln ala gly asp asp ser asn ser
1838/229
AAT AAG AAG AAT GCT GAC CTC CAA GTG CTA AAG CCC GAG CCC GAG CTG GTT TAT GAA GAC
asn lys lys asn ala asp leu gln val leu lys pro glu pro glu leu val tyr glu asp
1898/249
CTG AGG GGC TCA GTG ACC TTC CAC TGT GCC CTG GGC CCT GAG GTG GCA AAC GTG GCC AAA
leu arg gly ser val thr phe his cys ala leu gly pro glu val ala asn val ala lys
1958/269
TTT CTG TGC CGA CAG AGC AGT GGG GAA AAC TGT GAC GTG GTC GTC AAC ACC CTG GGG AAG
phe leu cys arg gln ser ser gly glu asn cys asp val val val asn thr leu gly lys
2018/289
AGG GCC CCA GCC TTT GAG GGC AGG ATC CTG CTC AAC CCC CAG GAC AAG GAT GGC TCA TTC
arg ala pro ala phe glu gly arg ile leu leu asn pro gln asp lys asp gly ser phe
2078/309
AGT GTG GTG ATC ACA GGC CTG AGG AAG GAG GAT GCA GGG CGC TAC CTG TGT GGA GCC CAT
ser val val ile thr gly leu arg lys glu asp ala gly arg tyr leu cys gly ala his
2138/329
TCG GAT GGT CAG CTG CAG GAA GGC TCG CCT ATC CAG GCC TGG CAA CTC TTC GTC AAT GAG
ser asp gly gln leu gln glu gly ser pro ile gln ala trp gln leu phe val asn glu
2198/349
GAG TCC ACG ATT CCC CGC AGC CCC ACT GTG GTG AAG GGG GTG GCA GGA AGC TCT GTG GCC
```

FIGURE 8 D – 1

glu ser thr ile pro arg ser pro thr val val lys gly val ala gly ser ser val ala
2258/369
GTG CTC TGC CCC TAC AAC CGT AAG GAA AGC AAA AGC ATC AAG TAC TGG TGT CTC TGG GAA
val leu cys pro tyr asn arg lys glu ser lys ser ile lys tyr trp cys leu trp glu
2318/389
GGG GCC CAG AAT GGC CGC TGC CCC CTG CTG GTG GAC AGC GAG GGG TGG GTT AAG GCC CAG
gly ala gln asn gly arg cys pro leu leu val asp ser glu gly trp val lys ala gln
2378/409
TAC GAG GGC CGC CTC TCC CTG CTG GAG GAG CCA GGC AAC GGC ACC TTC ACT GTC ATC CTC
tyr glu gly arg leu ser leu leu glu glu pro gly asn gly thr phe thr val ile leu
2438/429
AAC CAG CTC ACC AGC CGG GAC GCC GGC TTC TAC TGG TGT CTG ACC AAC GGC GAT ACT CTC
asn gln leu thr ser arg asp ala gly phe tyr trp cys leu thr asn gly asp thr leu
2498/449
TGG AGG ACC ACC GTG GAG ATC AAG ATT ATC GAA GGA GAA CCA AAC CTC AAG GTT CCC GGG
trp arg thr thr val glu ile lys ile ile glu gly glu pro asn leu lys val pro gly
2558/469
AAT GTC ACG GCT GTG CTG GGA GAG ACT CTC AAG GTC CCC TGT CAC TTT CCA TGC AAA TTC
asn val thr ala val leu gly glu thr leu lys val pro cys his phe pro cys lys phe
2618/489
TCC TCG TAC GAG AAA TAC TGG TGC AAG TGG AAT AAC ACG GGC TGC CAG GCC CTG CCC AGC
ser ser tyr glu lys tyr trp cys lys trp asn asn thr gly cys gln ala leu pro ser
2678/509
CAA GAC GAA GGC CCC AGC AAG GCC TTC GTG AAC TGT GAC GAG AAC AGC CGG CTT GTC TCC
gln asp glu gly pro ser lys ala phe val asn cys asp glu asn ser arg leu val ser
2738/529
CTG ACC CTG AAC CTG GTG ACC AGG GCT GAT GAG GGC TGG TAC TGG TGT GGA GTG AAG CAG
leu thr leu asn leu val thr arg ala asp glu gly trp tyr trp cys gly val lys gln
2798/549
GGC CAC TTC TAT GGA GAG ACT GCA GCC GTC TAT GTG GCA GTT GAA GAG AGG AAG GCA GCG
gly his phe tyr gly glu thr ala ala val tyr val ala val glu glu arg lys ala ala
2858/569
GGG TCC CGC GAT GTC AGC CTA GCG AAG GCA GAC GCT GCT CCT GAT GAG AAG GTG CTA GAC
gly ser arg asp val ser leu ala lys ala asp ala ala pro asp glu lys val leu asp
2918/589
TCT GGT TTT CGG GAG ATT GAG AAC AAA GCC ATT CAG GAT CCC AGG CTT TTT GCA GAG TGA
ser gly phe arg glu ile glu asn lys ala ile gln asp pro arg leu phe ala glu
2978
ATTC

FIGURE 8 D - 2

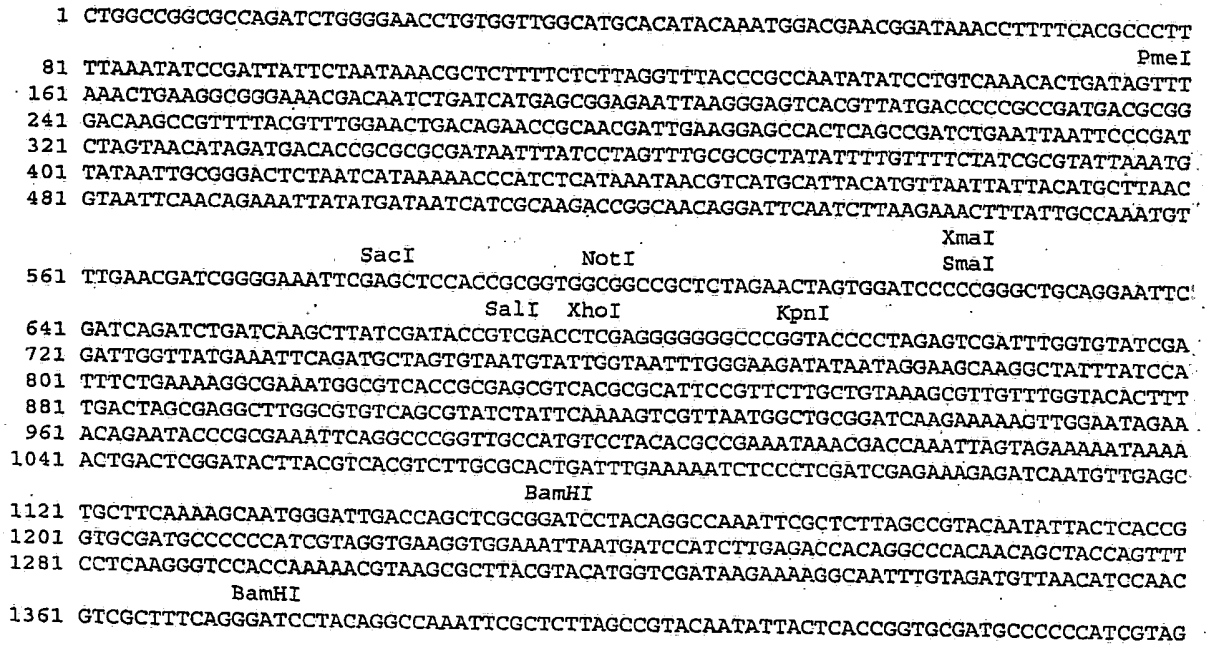

1 CTGGCCGGCGCCAGATCTGGGGAACCTGTGGTTGGCATGCACATACAAATGGACGAACGGATAAACCTTTTCACGCCCTT
PmeI
81 TTAAATATCCGATTATTCTAATAAACGCTCTTTTCTCTTAGGTTTACCCGCCAATATATCCTGTCAAACACTGATAGTTT
161 AAACTGAAGGCGGGAAACGACAATCTGATCATGAGCGGAGAATTAAGGGAGTCACGTTATGACCCCCGCCGATGACGCGG
241 GACAAGCCGTTTTACGTTTGGAACTGACAGAACCGCAACGATTGAAGGAGCCACTCAGCCGATCTGAATTAATTCCCGAT
321 CTAGTAACATAGATGACACCGCGCGCGATAATTTATCCTAGTTTGCGCGCTATATTTTGTTTTCTATCGCGTATTAAATG
401 TATAATTGCGGGACTCTAATCATAAAAACCCATCTCATAAATAACGTCATGCATTACATGTTAATTATTACATGCTTAAC
481 GTAATTCAACAGAAATTATATGATAATCATCGCAAGACCGGCAACAGGATTCAATCTTAAGAAACTTTATTGCCAAATGT
XmaI
SacI NotI SmaI
561 TTGAACGATCGGGGAAATTCGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTC
SalI XhoI KpnI
641 GATCAGATCTGATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCCTAGAGTCGATTTGGTGTATCGA
721 GATTGGTTATGAAATTCAGATGCTAGTGTAATGTATTGGTAATTTGGGAAGATATAATAGGAAGCAAGGCTATTTATCCA
801 TTTCTGAAAAGGCGAAATGGCGTCACCGCGAGCGTCACGCGCATTCCGTTCTTGCTGTAAAGCGTTGTTTGGTACACTTT
881 TGACTAGCGAGGCTTGGCGTGTCAGCGTATCTATTCAAAAGTCGTTAATGGCTGCGGATCAAGAAAAAGTTGGAATAGAA
961 ACAGAATACCCGCGAAATTCAGGCCCGGTTGCCATGTCCTACACGCCGAAATAAACGACCAAATTAGTAGAAAAATAAAA
1041 ACTGACTCGGATACTTACGTCACGTCTTGCGCACTGATTTGAAAAATCTCCCTCGATCGAGAAAGAGATCAATGTTGAGC
BamHI
1121 TGCTTCAAAAGCAATGGGATTGACCAGCTCGCGGATCCTACAGGCCAAATTCGCTCTTAGCCGTACAATATTACTCACCG
1201 GTGCGATGCCCCCCATCGTAGGTGAAGGTGGAAATTAATGATCCATCTTGAGACCACAGGCCCACAACAGCTACCAGTTT
1281 CCTCAAGGGTCCACCAAAAACGTAAGCGCTTACGTACATGGTCGATAAGAAAAGGCAATTTGTAGATGTTAACATCCAAC
BamHI
1361 GTCGCTTTCAGGGATCCTACAGGCCAAATTCGCTCTTAGCCGTACAATATTACTCACCGGTGCGATGCCCCCCATCGTAG

FIGURE 8 E – 1

```
1441 GTGAAGGTGGAAATTAATGATCCATCTTGAGACCACAGGCCCACAACAGCTACCAGTTTCCTCAAGGGTCCACCAAAAAC
                                                                          BamHI
1521 GTAAGCGCTTACGTACATGGTCGATAAGAAAAGGCAATTTGTAGATGTTAACATCCAACGTCGCTTTCAGGGATCCTACA
1601 GGCCAAATTCGCTCTTAGCCGTACAATATTACTCACCGGTGCGATGCCCCCCATCGTAGGTGAAGGTGGAAATTAATGAT
1681 CCATCTTGAGACCACAGGCCCACAACAGCTACCAGTTTCCTCAAGGGTCCACCAAAAACGTAAGCGCTTACGTACATGGT
                                                          BamHI
1761 CGATAAGAAAAGGCAATTTGTAGATGTTAACATCCAACGTCGCTTTCAGGGATCCGCGAGCTTATCGCGATACCGTCGAA
1841 TATAATAATTATATTTGTAAGAATATTATTATAATAATATAAAATATATATATATAAATTATAATATATTAATTATTGTT
1921 AATTATTAATAATATATATATTAATCATTTAGATATATAATTCTATAGCCTTAGACTCCTCATCAATAGAAGACTACGTA
2001 TAAAAATAATCAGATAACATCTAAAACATGTAGATAAATAATAGTTGTTTCATATCCAACATGATGTCCAGAGCTTCACG
2081 CTGCCGCAAGCACTCAGGGCGCAAGGGCTGCTAAAGGAAGCGGAACACGTAGAAAGCCAGTCCGCAGAANCGGTGCTGAC
2161 CCCGGATGAATGTCAGCTACTGGGCTATCTGGACAAGGGAAAACGCAAGCGCANNAGAGAAAGCAGGTAGCTTGCAGTGG
2241 GCTTACATGGCGATAGCTAGACTGGGCGGTTTTATGGACAGCAAGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTA
2321 AGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCATGA
2401 GCGGAGAATTAAGGGAGTCACGTTATGACCCCCGCCGATGACGCGGGACAAGCCGTTTTACGTTTGGAACTGACAGAACC
2481 GCAACGTTGAAGGAGCCACTCAGCCGCGGGTTTCTGGAGTTTAATGAGCTAAGCACATACGTCAGAAACCATTATTGCGC
2561 GTTCAAAAGTCGCCTAAGGTCACTATCAGCTAGCAAATATTTCTTGTCAAAAATGCTCCACTGACGTTCCATAAATTCCC
2641 CTCGGTATCCAATTAGAGTCTCATATTCACTCTCAATCCAGATCTGGATCGTTTCGCATGATTGAACAAGATGGATTGCA
                                                                MetIleGluGlnAspGlyLeuHi
2721 CGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCG
     sAlaGlySerProAlaAlaTrpValGluArgLeuPheGlyTyrAspTrpAlaGlnGlnThrIleGlyCysSerAspAlaA
2801 CCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAG
     laValPheArgLeuSerAlaGlnGlyArgProValLeuPheValLysThrAspLeuSerGlyAlaLeuAsnGluLeuGln
2881 GACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGG
     AspGluAlaAlaArgLeuSerTrpLeuAlaThrThrGlyValProCysAlaAlaValLeuAspValValThrGluAlaGl
2961 AAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCA
     yArgAspTrpLeuLeuLeuGlyGluValProGlyGlnAspLeuLeuSerSerHisLeuAlaProAlaGluLysValSerI
3041 TCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATC
     leMetAlaAspAlaMetArgArgLeuHisThrLeuAspProAlaThrCysProPheAspHisGlnAlaLysHisArgIle
3121 GAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGC
     GluArgAlaArgThrArgMetGluAlaGlyLeuValAspGlnAspAspLeuAspGluGluHisGlnGlyLeuAlaProAl
3201 CGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGATGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGA
     aGluLeuPheAlaArgLeuLysAlaArgMetProAspGlyAspAspLeuValValThrHisGlyAspAlaCysLeuProA
3281 ATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATA
     snIleMetValGluAsnGlyArgPheSerGlyPheIleAspCysGlyArgLeuGlyValAlaAspArgTyrGlnAspIle
3361 GCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGC
     AlaLeuAlaThrArgAspIleAlaGluGluLeuGlyGlyGluTrpAlaAspArgPheLeuValLeuTyrGlyIleAlaAl
                                                                          BamHI
3441 TCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGAGGATCCCCCGATGAGC
     aProAspSerGlnArgIleAlaPheTyrArgLeuLeuAspGluPhePhe∘∘∘
3521 TAAGCTAGCTATATCATCAATTTATGTATTACACATAATATCGCACTCAGTCTTTCATCTACGGCAATGTACCAGCTGAT
3601 ATAATCAGTTATTGAAATATTTCTGAATTTAAACTTGCATCAATAAATTTATGTTTTTGCTTGGACTATAATACCTGACT
3681 TGTTATTTTATCAATAAATATTTAAACTATATTTCTTTCAAGATGGGAATTAATTCACTGGCCGTCGTTTTACAACGTCG
3761 TGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAG
          PvuI                          BglI
3841 AGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCT
3921 CGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACC
4001 TCGACCCCAAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACG
4081 TTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGGCTATTCTTTTGA
4161 TTTATAAGGGATTTTGCCGATTTCGGAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTT
4241 GCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCCAG
4321 TACATTAAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACACCACAATATATCCTGCCACCAGC
4401 CAGCCAACAGCTCCCCGACCGGCAGCTCGGCACAAAATCACCACTCGATACAGGCAGCCCATCAG
```

FIGURE 8 E – 2

```
   1 CTGATGGGCTGCCTGTATCGAGTGGTGATTTTGTGCCGAGCTGCCGGTCGGGGAGCTGTTGGCTGGCTGGTGGCAGGATA
  81 TATTGTGGTGTAAACAAATTGACGCTTAGACAACTTAATAACACATTGCGGACGTTTTTAATGTACTGGGGTGGTTTTTC
 161 TTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTG
 241 GTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCC
 321 GAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAA
 401 AACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCAC
 481 TAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAG
                   Sfo I
 561 AAAGCGAAAGGAGCGGGCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCC
 641 AGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACG
           Xmn I
 721 ACGGCCAGTGAATTAATTCCCATCTTGAAAGAAATATAGTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAG
                                                 Xmn I
 801 TCCAAGCAAAAACATAAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGGTACATT
 881 GCCGTAGATGAAAGACTGAGTGCGATATTATGTGTAATACATAAATTGATGATATAGCTAGCTTAGCTCATCGGGGGATC
 961 CTCAGAGTCCCGCTCAGAAGAACTCGTCAAGAAGGCGATAGAAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGTAA
               ...PhePheGluAspLeuLeuArgTyrPheAlaIleArgGlnSerAspProAlaAlaIleGlyTyrL
1041 AGCACGAGGAAGCGGTCAGCCCATTCGCCGCCAAGCTCTTCAGCAATATCACGGGTAGCCAACGCTATGTCCTGATAGCG
     euValLeuPheArgAspAlaTrpGluGlyGlyLeuGluGluAlaIleAspArgThrAlaLeuAlaIleAspGlnTyrArg
1121 GTCCGCCACACCCAGCCGGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTTCCACCATGATATTCGGCAAGCAGGCAT
     AspAlaValGlyLeuArgGlyCysAspIlePheGlySerPheArgGlyAsnGluValMetIleAsnProLeuCysAlaAs
1201 CGCCATGGGTCACGACGAGATCATCGCCGTCGGGCATGCGCGCCTTGAGCCTGGCGAACAGTTCGGCTGGCGCGAGCCCC
     pGlyHisThrValValLeuAspAspGlyAspProMetArgAlaLysLeuArgAlaPheLeuGluAlaProAlaLeuGlyG
1281 TGATGCTCTTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCATCCGAGTACGTGCTCGCTCGATGCGATGTTTCGC
```

FIGURE 8 F – 1

```
    lnHisGluGluAspLeuAspAspGlnAspValLeuGlyAlaGluMetArgThrArgAlaArgGluIleArgHisLysAla
1361 TTGGTGGTCGAATGGGCAGGTAGCCGGATCAAGCGTATGCAGCCGCCGCATTGCATCAGCCATGATGGATACTTTCTCGG
    GlnHisAspPheProCysThrAlaProAspLeuThrHisLeuArgArgMetAlaAspAlaMetIleSerValLysGluAl
1441 CAGGAGCAAGGTGAGATGACAGGAGATCCTGCCCCGGCACTTCGCCCAATAGCAGCCAGTCCCTTCCCGCTTCAGTGACA
    aProAlaLeuHisSerSerLeuLeuAspGlnGlyProValGluGlyLeuLeuLeuTrpAspArgGlyAlaGluThrValV
                                                                    PstI
1521 ACGTCGAGCACAGCTGCGCAAGGAACGCCCGTCGTGGCCAGCCACGATAGCCGCGCTGCCTCGTCCTGCAGTTCATTCAG
    alAspLeuValAlaAlaCysProValGlyThrThrAlaLeuTrpSerLeuArgAlaAlaGluAspGlnLeuGluAsnLeu
                                        Sfo I
1601 GGCACCGGACAGGTCGGTCTTGACAAAAAGAACCGGGCGCCCCTGCGCTGACAGCCGGAACACGGCGGCATCAGAGCAGC
    AlaGlySerLeuAspThrLysValPheLeuValProArgGlyGlnAlaSerLeuArgPheValAlaAlaAspSerCysGl
1681 CGATTGTCTGTTGTGCCCAGTCATAGCCGAATAGCCTCTCCACCCAAGCGGCCGGAGAACCTGCGTGCAATCCATCTTGT
    yIleThrGlnGlnAlaTrpAspTyrGlyPheLeuArgGluValTrpAlaAlaProSerGlyAlaHisLeuGlyAspGlnG
                            BglII
1761 TCAATCATGCGAAACGATCCAGATCTGGATTGAGAGTGAATATGAGACTCTAATTGGATACCGAGGGGAATTTATGGAAC
    luIleMet
1841 GTCAGTGGAGCATTTTTGACAAGAAATATTTGCTAGCTGATAGTGACCTTAGGCGACTTTTGAACGCGCAATAATGGTTT
                                        SacII
1921 CTGACGTATGTGCTTAGCTCATTAAACTCCAGAAACCCGCGGCTGAGTGGCTCCTTCAACGTTGCGGTTCTGTCAGTTCC
2001 AAACGTAAAACGGCTTGTCCCGCGTCATCGGCGGGGGTCATAACGTGACTCCCTTAATTCTCCGCTCATGATCTTGATCC
                                                                    Sfo I
2081 CCTGCGCCATCAGATCCTTGGCGGCAAGAAAGCCATCCAGTTTACTTTGCAGGGCTTCCCAACCTTACCAGAGGGCGCCC
2161 CAGCTGGCAATTCCGGTTCGCTTGCTGTCCATAAAACCGCCCAGTCTAGCTATCGCCATGTAAGCCCACTGCAAGCTACC
2241 TGCTTTCTCTTTGCGCTTGCGTTTTCCCTTGTCCAGATAGCCCAGTAGCTGACATTCATCCGGGGTCAGCACCGNTTCTG
2321 CGGACTGGCTTTCTACGTGTTCCGCTTCCTTTAGCAGCCCTTGCGCCCTGAGTGCTTGCGGCAGCGTGAAGCTCTGGACA
2401 TCATGTTGGATATGAAACAACTATTATTTATCTACATGTTTTAGATGTTATCTGATTATTTTTATACGTAGTCTTCTATT
2481 GATGAGGAGTCTAAGGCTATAGAATTATATATCTAAATGATTAATATATATATTATTAATAATTAACAATAATTAATATA
                                                                    NruI
2561 TTATAATTTATATATATATATTTTATATTATTATAATAATATTCTTACAAATATAATTATTATATTCGACGGTATCGCGA
2641 TAAGCTCGCGGATCCCTGAAAGCGACGTTGGATGTTAACATCTACAAATTGCCTTTTCTTATCGACCATGTACGTAAGCG
2721 CTTACGTTTTTGGTGGACCCTTGAGGAAACTGGTAGCTGTTGTGGGCCTGTGGTCTCAAGATGGATCATTAATTTCCACC
2801 TTCACCTACGATGGGGGGCATCGCACCGGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCCTGTAGGATCCCTGAAA
2881 GCGACGTTGGATGTTAACATCTACAAATTGCCTTTTCTTATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGACCCT
2961 TGAGGAAACTGGTAGCTGTTGTGGGCCTGTGGTCTCAAGATGGATCATTAATTTCCACCTTCACCTACGATGGGGGGCAT
3041 CGCACCGGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCCTGTAGGATCCCTGAAAGCGACGTTGGATGTTAACATC
3121 TACAAATTGCCTTTTCTTATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGACCCTTGAGGAAACTGGTAGCTGTTG
3201 TGGGCCTGTGGTCTCAAGATGGATCATTAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGTAATATTGT
3281 ACGGCTAAGAGCGAATTTGGCCTGTAGGATCCGCGAGCTGGTCAATCCCATTGCTTTTGAAGCAGCTCAACATTGATCTC
3361 TTTCTCGATCGAGGGAGATTTTTCAAATCAGTGCGCAAGACGTGACGTAAGTATCCGAGTCAGTTTTTATTTTTCTACTA
3441 ATTTGGTCGTTTATTTCGGCGTGTAGGACATGGCAACCGGGCCTGAATTTCGCGGGTATTCTGTTTCTATTCCAACTTTT
3521 TCTTGATCCGCAGCCATTAACGACTTTTGAATAGATACGCTGACACGCCAAGCCTCGCTAGTCAAAAGTGTACCAAACAA
3601 CGCTTTACAGCAAGAACGGAATGCGCGTGACGCTCGCGGTGACGCCATTTCGCCTTTTCAGAAATGGATAAATAGCCTTG
3681 CTTCCTATTATATCTTCCCAAATTACCAATACATTACACTAGCATCTGAATTTCATAACCAATCTCGATACACCAAATCG
          KpnI
3761 ACTCTAGGGGTACCATGGTGCTCTTCGTGCTCACCTGCCTGCTGGCGGTCTTCCCAGCCATC
                    MetValLeuPheValLeuThrCysLeuLeuAlaValPheProAlaIle
3823 TCCACGAAGAGTCCCATATTTGGTCCCGAGGAGGTGAATAGTGTGGAAGGTAACTCAGTG
    SerThrLysSerProIlePheGlyProGluGluValAsnSerValGluGlyAsnSerVal
3883 TCCATCACGTGCTACTACCCACCCACCTCTGTCAACCGGCACACCCGGAAGTACTGGTGC
    SerIleThrCysTyrTyrProProThrSerValAsnArgHisThrArgLysTyrTrpCys
3943 CGGCAGGGAGCTAGAGGTGGCTGCATAACCCTCATCTCCTCGGAGGGCTACGTCTCCAGC
    ArgGlnGlyAlaArgGlyGlyCysIleThrLeuIleSerSerGluGlyTyrValSerSer
4003 AAATATGCAGGCAGGGCTAACCTCACCAACTTCCCGGAGAACGGCACATTTGTGGTGAAC
    LysTyrAlaGlyArgAlaAsnLeuThrAsnPheProGluAsnGlyThrPheValValAsn
4063 ATTGCCCAGCTGAGCCAGGATGACTCCGGGCGCTACAAGTGTGGCCTGGGCATCAATAGC
    IleAlaGlnLeuSerGlnAspAspSerGlyArgTyrLysCysGlyLeuGlyIleAsnSer
4123 CGAGGCCTGTCCTTTGATGTCAGCCTGGAGGTCAGCCAGGGTCCTGGGCTCCTAAATGAC
```

FIGURE 8 F – 2

```
     ArgGlyLeuSerPheAspValSerLeuGluValSerGlnGlyProGlyLeuLeuAsnAsp
4183 ACTAAAGTCTACACAGTGGACCTGGGCAGAACGGTGACCATCAACTGCCCTTTCAAGACT
     ThrLysValTyrThrValAspLeuGlyArgThrValThrIleAsnCysProPheLysThr
4243 GAGAATGCTCAAAAGAGGAAGTCCTTGTACAAGCAGATAGGCCTGTACCCTGTGCTGGTC
     GluAsnAlaGlnLysArgLysSerLeuTyrLysGlnIleGlyLeuTyrProValLeuVal
4303 ATCGACTCCAGTGGTTATGTGAATCCCAACTATACAGGAAGAATACGCCTTGATATTCAG
     IleAspSerSerGlyTyrValAsnProAsnTyrThrGlyArgIleArgLeuAspIleGln
4363 GGTACTGGCCAGTTACTGTTCAGCGTTGTCATCAACCAACTCAGGCTCAGCGATGCTGGG
     GlyThrGlyGlnLeuLeuPheSerValValIleAsnGlnLeuArgLeuSerAspAlaGly
4423 CAGTATCTCTGCCAGGCTGGGGATGATTCCAATAGTAATAAGAAGAATGCTGACCTCCAA
     GlnTyrLeuCysGlnAlaGlyAspAspSerAsnSerAsnLysLysAsnAlaAspLeuGln
4483 GTGCTAAAGCCCGAGCCCGAGCTGGTTTATGAAGACCTGAGGGGCTCAGTGACCTTCCAC
     ValLeuLysProGluProGluLeuValTyrGluAspLeuArgGlySerValThrPheHis
4543 TGTGCCCTGGGCCCTGAGGTGGCAAACGTGGCCAAATTTCTGTGCCGACAGAGCAGTGGG
     CysAlaLeuGlyProGluValAlaAsnValAlaLysPheLeuCysArgGlnSerSerGly
4603 GAAAACTGTGACGTGGTCGTCAACACCCTGGGGAAGAGGGCCCCAGCCTTTGAGGGCAGG
     GluAsnCysAspValValValAsnThrLeuGlyLysArgAlaProAlaPheGluGlyArg
4663 ATCCTGCTCAACCCCCAGGACAAGGATGGCTCATTCAGTGTGGTGATCACAGGCCTGAGG
     IleLeuLeuAsnProGlnAspLysAspGlySerPheSerValValIleThrGlyLeuArg
4723 AAGGAGGATGCAGGGCGCTACCTGTGTGGAGCCCATTCGGATGGTCAGCTGCAGGAAGGC
     LysGluAspAlaGlyArgTyrLeuCysGlyAlaHisSerAspGlyGlnLeuGlnGluGly
4783 TCGCCTATCCAGGCCTGGCAACTCTTCGTCAATGAGGAGTCCACGATTCCCCGCAGCCCC
     SerProIleGlnAlaTrpGlnLeuPheValAsnGluGluSerThrIleProArgSerPro
4843 ACTGTGGTGAAGGGGGTGGCAGGAAGCTCTGTGGCCGTGCTCTGCCCCTACAACCGTAAG
     ThrValValLysGlyValAlaGlySerSerValAlaValLeuCysProTyrAsnArgLys
4903 GAAAGCAAAAGCATCAAGTACTGGTGTCTCTGGGAAGGGGCCCAGAATGGCCGCTGCCCC
     GluSerLysSerIleLysTyrTrpCysLeuTrpGluGlyAlaGlnAsnGlyArgCysPro
4963 CTGCTGGTGGACAGCGAGGGGTGGGTTAAGGCCCAGTACGAGGGCCGCCTCTCCCTGCTG
     LeuLeuValAspSerGluGlyTrpValLysAlaGlnTyrGluGlyArgLeuSerLeuLeu
5023 GAGGAGCCAGGCAACGGCACCTTCACTGTCATCCTCAACCAGCTCACCAGCCGGGACGCC
     GluGluProGlyAsnGlyThrPheThrValIleLeuAsnGlnLeuThrSerArgAspAla
5083 GGCTTCTACTGGTGTCTGACCAACGGCGATACTCTCTGGAGGACCACCGTGGAGATCAAG
     GlyPheTyrTrpCysLeuThrAsnGlyAspThrLeuTrpArgThrThrValGluIleLys
5143 ATTATCGAAGGAGAACCAAACCTCAAGGTTCCCGGGAATGTCACGGCTGTGCTGGGAGAG
     IleIleGluGlyGluProAsnLeuLysValProGlyAsnValThrAlaValLeuGlyGlu
5203 ACTCTCAAGGTCCCCTGTCACTTTCCATGCAAATTCTCCTCGTACGAGAAATACTGGTGC
     ThrLeuLysValProCysHisPheProCysLysPheSerSerTyrGluLysTyrTrpCys
5263 AAGTGGAATAACACGGGCTGCCAGGCCCTGCCCAGCCAAGACGAAGGCCCCAGCAAGGCC
     LysTrpAsnAsnThrGlyCysGlnAlaLeuProSerGlnAspGluGlyProSerLysAla
5323 TTCGTGAACTGTGACGAGAACAGCCGGCTTGTCTCCCTGACCCTGAACCTGGTGACCAGG
     PheValAsnCysAspGluAsnSerArgLeuValSerLeuThrLeuAsnLeuValThrArg
5383 GCTGATGAGGGCTGGTACTGGTGTGGAGTGAAGCAGGGCCACTTCTATGGAGAGACTGCA
     AlaAspGluGlyTrpTyrTrpCysGlyValLysGlnGlyHisPheTyrGlyGluThrAla
5443 GCCGTCTATGTGGCAGTTGAAGAGAGGAAGGCAGCGGGGTCCCGCGATGTCAGCCTAGCG
     AlaValTyrValAlaValGluGluArgLysAlaAlaGlySerArgAspValSerLeuAla
5503 AAGGCAGACGCTGCTCCTGATGAGAAGGTGCTAGACTCTGGTTTTCGGGAGATTGAGAAC
     LysAlaAspAlaAlaProAspGluLysValLeuAspSerGlyPheArgGluIleGluAsn
                                    EcoRI
5563 AAAGCCATTCAGGATCCCAGGCTTTTTGCAGAGTGAATTCCCGATCGTTCAAACATTTGGCAATAAAG
     LysAlaIleGlnAspProArgLeuPheAlaGlu
5631 TTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAA
5711 TTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATA
5791 GAAAACAAAATATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGGGGATCCGTCGA
        ClaI
5871 CGGTATCGATAAGGATCCCTGAAAGCGACGTTGGATGTTAACATCTACAAATTGCCTTTTCTTATCGACCATGTACGTAA
5951 GCGCTTACGTTTTTGGTGGACCCTTGAGGAAACTGGTAGCTGTTGTGGGCCTGTGGTCTCAAGATGGATCATTAATTTCC
6031 ACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCCTGTAGGATCCCTG
```

FIGURE 8 F – 3

```
6111 AAAGCGACGTTGGATGTTAACATCTACAAATTGCCTTTTCTTATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGAC
6191 CCTTGAGGAAACTGGTAGCTGTTGTGGGCCTGTGGTCTCAAGATGGATCATTAATTTCCACCTTCACCTACGATGGGGGG
6271 CATCGCACCGGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCCTGTAGGATCCCTGAAAGCGACGTTGGATGTTAAC
6351 ATCTACAAATTGCCTTTTCTTATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGACCCTTGAGGAAACTGGTAGCTG
6431 TTGTGGGCCTGTGGTCTCAAGATGGATCATTAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGTAATAT
6511 TGTACGGCTAAGAGCGAATTTGGCCTGTAGGATCCGCGAGCTGGTCAATCCCATTGCTTTTGAAGCAGCTCAACATTGAT
          XhoI
6591 CTCTTTCTCGAGGGAGATTTTTCAAATCAGTGCGCAAGACGTGACGTAAGTATCCGAGTCAGTTTTTATTTTTCTACTAA
6671 TTTGGTCGTTTATTTCGGCGTGTAGGACATGGCAACCGGGCCTGAATTTCGCGGGTATTCTGTTTCTATTCCAACTTTTT
6751 CTTGATCCGCAGCCATTAACGACTTTTGAATAGATACGCTGACACGCCAAGCCTCGCTAGTCAAAAGTGTACCAAACAAC
6831 GCTTTACAGCAAGAACGGAATGCGCGTGACGCTCGCGGTGACGCCATTTCGCCTTTTCAGAAATGGATAAATAGCCTTGC
6911 TTCCTATTATATCTTCCCAAATTACCAATACATTACACTAGCATCTGAATTTCATAACCAATCTCGATACACCAAATCGA
      XbaI
6991 CTCTAGAGGATCTAACCATGGGATCTAAACCTTTTTTGTCTCTTCTTTCATTGTCATTGCTT
                        MetGlySerLysProPheLeuSerLeuLeuSerLeuSerLeuLeu
                   SpeI
7053 TTGTTTACATCTACTAGTTTGGCACAAGAAGATGAAAGGATTGTTCTTGTTGACAACAAA
     LeuPheThrSerThrSerLeuAlaGlnGluAspGluArgIleValLeuValAspAsnLys
7113 TGTAAGTGTGCCCGGATTACTTCCAGGATCATCCGTTCTTCCGAAGATCCTAATGAGGAC
     CysLysCysAlaArgIleThrSerArgIleIleArgSerSerGluAspProAsnGluAsp
7173 ATTGTGGAGAGAAACATCCGAATTATTGTTCCTCTGAACAACAGGGAGAATATCTCTGAT
     IleValGluArgAsnIleArgIleIleValProLeuAsnAsnArgGluAsnIleSerAsp
7233 CCCACCTCACCATTGAGAACCAGATTTGTGTACCATTTGTCTGACCTCTGTAAAAAATGT
     ProThrSerProLeuArgThrArgPheValTyrHisLeuSerAspLeuCysLysLysCys
7293 GATCCTACAGAAGTGGAGCTGGATAATCAGATAGTTACTGCTACCCAGAGCAATATCTGT
     AspProThrGluValGluLeuAspAsnGlnIleValThrAlaThrGlnSerAsnIleCys
7353 GATGAAGACAGTGCTACAGAGACCTGCTACACTTATGACAGAAACAAGTGCTACACAGCT
     AspGluAspSerAlaThrGluThrCysTyrThrTyrAspArgAsnLysCysTyrThrAla
7413 GTGGTCCCACTCGTATATGGTGGTGAGACCAAAATGGTGGAAACAGCCTTAACCCCAGAT
     ValValProLeuValTyrGlyGlyGluThrLysMetValGluThrAlaLeuThrProAsp
                        SacI
7473 GCCTGCTATCCTGACTGAGCTCGAATTTCCCCGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAAT
     AlaCysTyrProAsp•••
7547 CCTGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCAT
7627 GACGTTATTTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGC
7707 GCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGGGAATTAATTCAGATCGGCTGAGTGGCT
7787 CCTTCAATCGTTGCGGTTCTGTCAGTTCCAAACGTAAAACGGCTTGTCCCGCGTCATCGGCGGGGGTCATAACGTGACTC
7867 CCTTAATTCTCCGCTCATGATCAGATTGTCGTTTCCCGCCTTCAGTTTAAACTATCAGTGTTTGACAGGATATATTGGCG
7947 GGTAAACCTAAGAGAAAAGAGCGTTTATTAGAATAATCGGATATTTAAAAGGGCGTGAAAAGGTTTATCCGTTCGTCCAT
                                   BglII Sfo I
8027 TTGTATGTGCATGCCAACCACAGGTTCCCCAGATCTGGCGCCGGCCAG
```

FIGURE 8 F – 4

FIGURE 9

SEQUENCE LISTING

<110> PLANET BIOTECHNOLOGY, INCORPORATED
Larrick, James William
Wycoff, Keith Lynn

<120> NOVEL IMMUNOADHESION FOR THE PREVENTION OF RHINOVIRUS INFECTION

<130> 030905.0003.WO

<140> To be assigned
<141> 2001-04-27

<150> 60/200,298
<151> 2000-04-27

<160> 8

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 1596
<212> DNA
<213> Homo Sapien

<220>
<221> CDS
<222> (1)...(1596)
<223> ICAM-1

<400> 1

```
atg gct ccc agc agc ccc cgg ccc gcg ctg ccc gca ctc ctg gtc ctg       48
Met Ala Pro Ser Ser Pro Arg Pro Ala Leu Pro Ala Leu Leu Val Leu
 1               5                  10                  15

ctc ggg gct ctg ttc cca gga cct ggc aat gcc cag aca tct gtg tcc       96
Leu Gly Ala Leu Phe Pro Gly Pro Gly Asn Ala Gln Thr Ser Val Ser
            20                  25                  30

ccc tca aaa gtc atc ctg ccc cgg gga ggc tcc gtg ctg gtg aca tgc      144
Pro Ser Lys Val Ile Leu Pro Arg Gly Gly Ser Val Leu Val Thr Cys
         35                  40                  45

agc acc tcc tgt gac cag ccc aag ttg ttg ggc ata gag acc ccg ttg      192
Ser Thr Ser Cys Asp Gln Pro Lys Leu Leu Gly Ile Glu Thr Pro Leu
     50                  55                  60

cct aaa aag gag ttg ctc ctg cct ggg aac aac cgg aag gtg tat gaa      240
Pro Lys Lys Glu Leu Leu Leu Pro Gly Asn Asn Arg Lys Val Tyr Glu
 65                  70                  75                  80

ctg agc aat gtg caa gaa gat agc caa cca atg tgc tat tca aac tgc      288
Leu Ser Asn Val Gln Glu Asp Ser Gln Pro Met Cys Tyr Ser Asn Cys
                85                  90                  95
```

FIGURE 9 – 1

```
cct gat ggg cag tca aca gct aaa acc ttc ctc acc gtg tac tgg act      336
Pro Asp Gly Gln Ser Thr Ala Lys Thr Phe Leu Thr Val Tyr Trp Thr
            100                 105                 110

cca gaa cgg gtg gaa ctg gca ccc ctc ccc tct tgg cag cca gtg ggc      384
Pro Glu Arg Val Glu Leu Ala Pro Leu Pro Ser Trp Gln Pro Val Gly
        115                 120                 125

aag aac ctt acc cta cgc tgc cag gtg gag ggt ggg gca ccc cgg gcc      432
Lys Asn Leu Thr Leu Arg Cys Gln Val Glu Gly Gly Ala Pro Arg Ala
    130                 135                 140

aac ctc acc gtg gtg ctg ctc cgt ggg gag aag gag ctg aaa cgg gag      480
Asn Leu Thr Val Val Leu Leu Arg Gly Glu Lys Glu Leu Lys Arg Glu
145                 150                 155                 160

cca gct gtg ggg gag ccc gct gag gtc acg acc acg gtg ctg gtg agg      528
Pro Ala Val Gly Glu Pro Ala Glu Val Thr Thr Thr Val Leu Val Arg
                165                 170                 175

aga gat cac cat gga gcc aat ttc tcg tgc cgc act gaa ctg gac ctg      576
Arg Asp His His Gly Ala Asn Phe Ser Cys Arg Thr Glu Leu Asp Leu
            180                 185                 190

cgg ccc caa ggg ctg gag ctg ttt gag aac acc tcg gcc ccc tac cag      624
Arg Pro Gln Gly Leu Glu Leu Phe Glu Asn Thr Ser Ala Pro Tyr Gln
        195                 200                 205

ctc cag acc ttt gtc ctg cca gcg act ccc cca caa ctt gtc agc ccc      672
Leu Gln Thr Phe Val Leu Pro Ala Thr Pro Pro Gln Leu Val Ser Pro
    210                 215                 220

cgg gtc cta gag gtg gac acg cag ggg acc gtg gtc tgt tcc ctg gac      720
Arg Val Leu Glu Val Asp Thr Gln Gly Thr Val Val Cys Ser Leu Asp
225                 230                 235                 240

ggg ctg ttc cca gtc tcg gag gcc cag gtc cac ctg gca ctg ggg gac      768
Gly Leu Phe Pro Val Ser Glu Ala Gln Val His Leu Ala Leu Gly Asp
                245                 250                 255

cag agg ttg aac ccc aca gtc acc tat ggc aac gac tcc ttc tcg gcc      816
Gln Arg Leu Asn Pro Thr Val Thr Tyr Gly Asn Asp Ser Phe Ser Ala
            260                 265                 270

aag gcc tca gtc agt gtg acc gca gag gac gag ggc acc cag cgg ctg      864
Lys Ala Ser Val Ser Val Thr Ala Glu Asp Glu Gly Thr Gln Arg Leu
        275                 280                 285

acg tgt gca gta ata ctg ggg aac cag agc cag gag aca ctg cag aca      912
Thr Cys Ala Val Ile Leu Gly Asn Gln Ser Gln Glu Thr Leu Gln Thr
    290                 295                 300

gtg acc atc tac agc ttt ccg gcg ccc aac gtg att ctg acg aag cca      960
Val Thr Ile Tyr Ser Phe Pro Ala Pro Asn Val Ile Leu Thr Lys Pro
305                 310                 315                 320
```

FIGURE 9 – 2

```
gag gtc tca gaa ggg acc gag gtg aca gtg aag tgt gag gcc cac cct     1008
Glu Val Ser Glu Gly Thr Glu Val Thr Val Lys Cys Glu Ala His Pro
                325                 330                 335

aga gcc aag gtg acg ctg aat ggg gtt cca gcc cag cca ctg ggc ccg     1056
Arg Ala Lys Val Thr Leu Asn Gly Val Pro Ala Gln Pro Leu Gly Pro
            340                 345                 350

agg gcc cag ctc ctg ctg aag gcc acc cca gag gac aac ggg cgc agc     1104
Arg Ala Gln Leu Leu Leu Lys Ala Thr Pro Glu Asp Asn Gly Arg Ser
        355                 360                 365

ttc tcc tgc tct gca acc ctg gag gtg gcc ggc cag ctt ata cac aag     1152
Phe Ser Cys Ser Ala Thr Leu Glu Val Ala Gly Gln Leu Ile His Lys
    370                 375                 380

aac cag acc cgg gag ctt cgt gtc ctg tat ggc ccc cga ctg gac gag     1200
Asn Gln Thr Arg Glu Leu Arg Val Leu Tyr Gly Pro Arg Leu Asp Glu
385                 390                 395                 400

agg gat tgt ccg gga aac tgg acg tgg cca gaa aat tcc cag cag act     1248
Arg Asp Cys Pro Gly Asn Trp Thr Trp Pro Glu Asn Ser Gln Gln Thr
                405                 410                 415

cca atg tgc cag gct tgg ggg aac cca ttg ccc gag ctc aag tgt cta     1296
Pro Met Cys Gln Ala Trp Gly Asn Pro Leu Pro Glu Leu Lys Cys Leu
            420                 425                 430

aag gat ggc act ttc cca ctg ccc atc ggg gaa tca gtg act gtc act     1344
Lys Asp Gly Thr Phe Pro Leu Pro Ile Gly Glu Ser Val Thr Val Thr
        435                 440                 445

cga gat ctt gag ggc acc tac ctc tgt cgg gcc agg agc act caa ggg     1392
Arg Asp Leu Glu Gly Thr Tyr Leu Cys Arg Ala Arg Ser Thr Gln Gly
    450                 455                 460

gag gtc acc cgc aag gtg acc gtg aat gtg ctc tcc ccc cgg tat gag     1440
Glu Val Thr Arg Lys Val Thr Val Asn Val Leu Ser Pro Arg Tyr Glu
465                 470                 475                 480

att gtc atc atc act gtg gta gca gcc gca gtc ata atg ggc act gca     1488
Ile Val Ile Ile Thr Val Val Ala Ala Ala Val Ile Met Gly Thr Ala
                485                 490                 495

ggc ctc agc acg tac ctc tat aac cgc cag cgg aag atc aag aaa tac     1536
Gly Leu Ser Thr Tyr Leu Tyr Asn Arg Gln Arg Lys Ile Lys Lys Tyr
            500                 505                 510

aga cta caa cag gcc caa aaa ggg acc ccc atg aaa ccg aac aca caa     1584
Arg Leu Gln Gln Ala Gln Lys Gly Thr Pro Met Lys Pro Asn Thr Gln
        515                 520                 525

gcc acg cct ccc                                                    1596
Ala Thr Pro Pro
    530
```

FIGURE 9 – 3

<210> 2
<211> 532
<212> PRT
<213> Homo Sapien

<400> 2

```
Met Ala Pro Ser Ser Pro Arg Pro Ala Leu Pro Ala Leu Leu Val Leu
 1               5                  10                  15
Leu Gly Ala Leu Phe Pro Gly Pro Gly Asn Ala Gln Thr Ser Val Ser
            20                  25                  30
Pro Ser Lys Val Ile Leu Pro Arg Gly Gly Ser Val Leu Val Thr Cys
        35                  40                  45
Ser Thr Ser Cys Asp Gln Pro Lys Leu Leu Gly Ile Glu Thr Pro Leu
    50                  55                  60
Pro Lys Lys Glu Leu Leu Leu Pro Gly Asn Asn Arg Lys Val Tyr Glu
65                  70                  75                  80
Leu Ser Asn Val Gln Glu Asp Ser Gln Pro Met Cys Tyr Ser Asn Cys
                85                  90                  95
Pro Asp Gly Gln Ser Thr Ala Lys Thr Phe Leu Thr Val Tyr Trp Thr
            100                 105                 110
Pro Glu Arg Val Glu Leu Ala Pro Leu Pro Ser Trp Gln Pro Val Gly
        115                 120                 125
Lys Asn Leu Thr Leu Arg Cys Gln Val Glu Gly Gly Ala Pro Arg Ala
    130                 135                 140
Asn Leu Thr Val Val Leu Leu Arg Gly Glu Lys Glu Leu Lys Arg Glu
145                 150                 155                 160
Pro Ala Val Gly Glu Pro Ala Glu Val Thr Thr Thr Val Leu Val Arg
                165                 170                 175
Arg Asp His His Gly Ala Asn Phe Ser Cys Arg Thr Glu Leu Asp Leu
            180                 185                 190
Arg Pro Gln Gly Leu Glu Leu Phe Glu Asn Thr Ser Ala Pro Tyr Gln
        195                 200                 205
Leu Gln Thr Phe Val Leu Pro Ala Thr Pro Pro Gln Leu Val Ser Pro
    210                 215                 220
Arg Val Leu Glu Val Asp Thr Gln Gly Thr Val Val Cys Ser Leu Asp
225                 230                 235                 240
Gly Leu Phe Pro Val Ser Glu Ala Gln Val His Leu Ala Leu Gly Asp
                245                 250                 255
Gln Arg Leu Asn Pro Thr Val Thr Tyr Gly Asn Asp Ser Phe Ser Ala
            260                 265                 270
Lys Ala Ser Val Ser Val Thr Ala Glu Asp Glu Gly Thr Gln Arg Leu
        275                 280                 285
Thr Cys Ala Val Ile Leu Gly Asn Gln Ser Gln Glu Thr Leu Gln Thr
    290                 295                 300
Val Thr Ile Tyr Ser Phe Pro Ala Pro Asn Val Ile Leu Thr Lys Pro
305                 310                 315                 320
Glu Val Ser Glu Gly Thr Glu Val Thr Val Lys Cys Glu Ala His Pro
                325                 330                 335
Arg Ala Lys Val Thr Leu Asn Gly Val Pro Ala Gln Pro Leu Gly Pro
            340                 345                 350
Arg Ala Gln Leu Leu Leu Lys Ala Thr Pro Glu Asp Asn Gly Arg Ser
        355                 360                 365
Phe Ser Cys Ser Ala Thr Leu Glu Val Ala Gly Gln Leu Ile His Lys
    370                 375                 380
Asn Gln Thr Arg Glu Leu Arg Val Leu Tyr Gly Pro Arg Leu Asp Glu
385                 390                 395                 400
Arg Asp Cys Pro Gly Asn Trp Thr Trp Pro Glu Asn Ser Gln Gln Thr
```

FIGURE 9 – 4

```
                405                 410                 415
Pro Met Cys Gln Ala Trp Gly Asn Pro Leu Pro Glu Leu Lys Cys Leu
            420                 425                 430
Lys Asp Gly Thr Phe Pro Leu Pro Ile Gly Glu Ser Val Thr Val Thr
        435                 440                 445
Arg Asp Leu Glu Gly Thr Tyr Leu Cys Arg Ala Arg Ser Thr Gln Gly
    450                 455                 460
Glu Val Thr Arg Lys Val Thr Val Asn Val Leu Ser Pro Arg Tyr Glu
465                 470                 475                 480
Ile Val Ile Ile Thr Val Val Ala Ala Ala Val Ile Met Gly Thr Ala
                485                 490                 495
Gly Leu Ser Thr Tyr Leu Tyr Asn Arg Gln Arg Lys Ile Lys Lys Tyr
            500                 505                 510
Arg Leu Gln Gln Ala Gln Lys Gly Thr Pro Met Lys Pro Asn Thr Gln
        515                 520                 525
Ala Thr Pro Pro
    530
```

<210> 3
<211> 3003
<212> DNA
<213> Homo Sapien

<400> 3

```
gctataagga tcacgcgccc cagtcgacgc tgagctcctc tgctactcag agttgcaacc      60
tcagcctcgc tatggctccc agcagccccc ggcccgcgct gcccgcactc ctggtcctgc     120
tcggggctct gttcccagga cctggcaatg cccagacatc tgtgtccccc tcaaaagtca     180
tcctgccccg gggaggctcc gtgctggtga catgcagcac ctcctgtgac cagcccaagt     240
tgttgggcat agagaccccg ttgcctaaaa aggagttgct cctgcctggg aacaaccgga     300
aggtgtatga actgagcaat gtgcaagaag atagccaacc aatgtgctat tcaaactgcc     360
ctgatgggca gtcaacagct aaaaccttcc tcaccgtgta ctggactcca gaacgggtgg     420
aactggcacc cctcccctct tggcagccag tgggcaagaa ccttacccta cgctgccagg     480
tggagggtgg ggcaccccgg gccaacctca ccgtggtgct gctccgtggg gagaaggagc     540
tgaaacggga gccagctgtg ggggagcccg ctgaggtcac gaccacggtg ctggtgagga     600
gagatcacca tggagccaat ttctcgtgcc gcactgaact ggacctgcgg ccccaagggc     660
tggagctgtt tgagaacacc tcggcccctt accagctcca gacctttgtc ctgccagcga     720
ctcccccaca acttgtcagc ccccgggtcc tagaggtgga cacgcagggg accgtggtct     780
gttccctgga cgggctgttc ccagtctcgg aggcccaggt ccacctggca ctgggggacc     840
agaggttgaa ccccacagtc acctatggca acgactcctt ctcggccaag gcctcagtca     900
gtgtgaccgc agaggacgag ggcacccagc ggctgacgtg tgcagtaata ctggggaacc     960
agagccagga gacactgcag acagtgacca tctacagctt tccggcgccc aacgtgattc    1020
tgacgaagcc agaggtctca gaagggaccg aggtgacagt gaagtgtgag gcccacccta    1080
gagccaaggt gacgctgaat ggggttccag cccagccact gggcccgagg gcccagctcc    1140
tgctgaaggc caccccagag gacaacgggc gcagcttctc ctgctctgca accctggagg    1200
tggccggcca gcttatacac aagaaccaga cccgggagct tcgtgtcctg tatggccccc    1260
gactggacga gagggattgt ccgggaaact ggacgtggcc agaaaattcc cagcagactc    1320
caatgtgcca ggcttggggg aacccattgc ccgagctcaa gtgtctaaag gatggcactt    1380
tcccactgcc catcggggaa tcagtgactg tcactcgaga tcttgagggc acctacctct    1440
gtcgggccag gagcactcaa ggggaggtca cccgcaaggt gaccgtgaat gtgctctccc    1500
cccggtatga gattgtcatc atcactgtgg tagcagccgc agtcataatg ggcactgcag    1560
gcctcagcac gtacctctat aaccgccagc ggaagatcaa gaaatacaga ctacaacagg    1620
cccaaaaagg gacccccatg aaaccgaaca cacaagccac gcctccctga acctatcccg    1680
ggacagggcc tcttcctcgg ccttcccata ttggtggcag tggtgccaca ctgaacagag    1740
tggaagacat atgccatgca gctacaccta ccggccctgg gacgccggag gacagggcat    1800
tgtcctcagt cagatacaac agcatttggg gccatggtac ctgcacacct aaaacactag    1860
gccacgcatc tgatctgtag tcacatgact aagccaagag gaaggagcaa gactcaagac    1920
atgattgatg gatgttaaag tctagcctga tgagagggga agtggtgggg gagacatagc    1980
```

FIGURE 9 – 5

```
cccaccatga ggacatacaa ctgggaaata ctgaaacttg ctgcctattg ggtatgctga    2040
ggccccacag acttacagaa gaagtggccc tccatagaca tgtgtagcat caaaacacaa    2100
aggcccacac ttcctgacgg atgccagctt gggcactgct gtctactgac cccaaccctt    2160
gatgatatgt atttattcat ttgttatttt accagctatt tattgagtgt cttttatgta    2220
ggctaaatga acataggtct ctggcctcac ggagctccca gtccatgtca cattcaaggt    2280
caccaggtac agttgtacag gttgtacact gcaggagagt gcctggcaaa aagatcaaat    2340
ggggctggga cttctcattg gccaacctgc ctttccccag aaggagtgat ttttctatcg    2400
gcacaaaagc actatatgga ctggtaatgg ttcacaggtt cagagattac ccagtgaggc    2460
cttattcctc ccttcccccc aaaactgaca cctttgttag ccacctcccc acccacatac    2520
atttctgcca gtgttcacaa tgacactcag cggtcatgtc tggacatgag tgcccaggga    2580
atatgcccaa gctatgcctt gtcctcttgt cctgtttgca tttcactggg agcttgcact    2640
attgcagctc cagtttcctg cagtgatcag ggtcctgcaa gcagtgggga agggggccaa    2700
ggtattggag gactccctcc cagctttgga agcctcatcc gcgtgtgtgt gtgtgtgtgt    2760
atgtgtagac aagctctcgc tctgtcaccc aggctggagt gcagtggtgc aatcatggtt    2820
cactgcagtc ttgacctttt gggctcaagt gatcctccca cctcagcctc ctgagtagct    2880
gggaccatag gctcacaaca ccacacctgg caaatttgat ttttttttt tttttcagag    2940
acggggtctc gcaacattgc ccagacttcc tttgtgttag ttaataaagc tttctcaact    3000
gcc                                                                 3003
```

<210> 4
<211> 6
<212> PRT
<213> Homo Sapien

<400> 4
Ser Glu Lys Asp Glu Leu
1 5

<210> 5
<211> 7
<212> PRT
<213> Homo Sapien

<400> 5
Arg Ser Glu Lys Asp Glu Leu
1 5

<210> 6
<211> 52
<212> DNA
<213> Homo Sapien

<400> 6
tctgttccca ggaactagtt tggcacagac atctgtgtcc ccctcaaaag tc 52

<210> 7
<211> 38
<212> DNA
<213> Homo Sapien

<400> 7
cataccgggg actagtcaca ttcacggtca cctcgcgg 38

<210> 8
<211> 799
<212> PRT
<213> Homo Sapien

FIGURE 9 – 6

<220>
<221> CDS
<222> (1)...(448)
<223> ICAM-1 Extracellular Domains

<220>
<221> CDS
<222> (453)...(799)
<223> Human IgA2m(2)

<400> 8

```
Gln Thr Ser Val Ser Pro Ser Lys Val Ile Leu Pro Arg Gly Gly Ser
 1               5                  10                  15
Val Leu Val Thr Cys Ser Thr Ser Cys Asp Gln Pro Lys Leu Leu Gly
            20                  25                  30
Ile Glu Thr Pro Leu Pro Lys Lys Glu Leu Leu Leu Pro Gly Asn Asn
        35                  40                  45
Arg Lys Val Tyr Glu Leu Ser Asn Val Gln Glu Asp Ser Gln Pro Met
    50                  55                  60
Cys Tyr Ser Asn Cys Pro Asp Gly Gln Ser Thr Ala Lys Thr Phe Leu
65                  70                  75                  80
Thr Val Tyr Trp Thr Pro Glu Arg Val Glu Leu Ala Pro Leu Pro Ser
                85                  90                  95
Trp Gln Pro Val Gly Lys Asn Leu Thr Leu Arg Cys Gln Val Glu Gly
            100                 105                 110
Gly Ala Pro Arg Ala Asn Leu Thr Val Val Leu Leu Arg Gly Glu Lys
        115                 120                 125
Glu Leu Lys Arg Glu Pro Ala Val Gly Glu Pro Ala Glu Val Thr Thr
    130                 135                 140
Thr Val Leu Val Arg Arg Asp His His Gly Ala Asn Phe Ser Cys Arg
145                 150                 155                 160
Thr Glu Leu Asp Leu Arg Pro Gln Gly Leu Glu Leu Phe Glu Asn Thr
                165                 170                 175
Ser Ala Pro Tyr Gln Leu Gln Thr Phe Val Leu Pro Ala Thr Pro Pro
            180                 185                 190
Gln Leu Val Ser Pro Arg Val Leu Glu Val Asp Thr Gln Gly Thr Val
        195                 200                 205
Val Cys Ser Leu Asp Gly Leu Phe Pro Val Ser Glu Ala Gln Val His
    210                 215                 220
Leu Ala Leu Gly Asp Gln Arg Leu Asn Pro Thr Val Thr Tyr Gly Asn
225                 230                 235                 240
Asp Ser Phe Ser Ala Lys Ala Ser Val Ser Val Thr Ala Glu Asp Glu
                245                 250                 255
Gly Thr Gln Arg Leu Thr Cys Ala Val Ile Leu Gly Asn Gln Ser Gln
            260                 265                 270
Glu Thr Leu Gln Thr Val Thr Ile Tyr Ser Phe Pro Ala Pro Asn Val
        275                 280                 285
Ile Leu Thr Lys Pro Glu Val Ser Glu Gly Thr Glu Val Thr Val Lys
    290                 295                 300
Cys Glu Ala His Pro Arg Ala Lys Val Thr Leu Asn Gly Val Pro Ala
305                 310                 315                 320
Gln Pro Leu Gly Pro Arg Ala Gln Leu Leu Leu Lys Ala Thr Pro Glu
                325                 330                 335
Asp Asn Gly Arg Ser Phe Ser Cys Ser Ala Thr Leu Glu Val Ala Gly
            340                 345                 350
Gln Leu Ile His Lys Asn Gln Thr Arg Glu Leu Arg Val Leu Tyr Gly
        355                 360                 365
```

FIGURE 9 – 7

```
Pro Arg Leu Asp Glu Arg Asp Cys Pro Gly Asn Trp Thr Trp Pro Glu
    370                 375                 380
Asn Ser Gln Gln Thr Pro Met Cys Gln Ala Trp Gly Asn Pro Leu Pro
385                 390                 395                 400
Glu Leu Lys Cys Leu Lys Asp Gly Thr Phe Pro Leu Pro Ile Gly Glu
                405                 410                 415
Ser Val Thr Val Thr Arg Asp Leu Glu Gly Thr Tyr Leu Cys Arg Ala
            420                 425                 430
Arg Ser Thr Gln Gly Glu Val Thr Arg Glu Val Thr Val Asn Val Thr
        435                 440                 445
Ser Gly Ser Ser Ala Ser Pro Thr Ser Pro Lys Val Phe Pro Leu Ser
    450                 455                 460
Leu Asp Ser Thr Pro Gln Asp Gly Asn Val Val Val Ala Cys Leu Val
465                 470                 475                 480
Gln Gly Phe Phe Pro Gln Glu Pro Leu Ser Val Thr Trp Ser Glu Ser
                485                 490                 495
Gly Gln Asn Val Thr Ala Arg Asn Phe Pro Pro Ser Gln Asp Ala Ser
            500                 505                 510
Gly Asp Leu Tyr Thr Thr Ser Ser Gln Leu Thr Leu Pro Ala Thr Gln
        515                 520                 525
Cys Pro Asp Gly Lys Ser Val Thr Cys His Val Lys His Tyr Thr Asn
    530                 535                 540
Ser Ser Gln Asp Val Thr Val Pro Cys Arg Val Pro Pro Pro Pro Pro
545                 550                 555                 560
Cys Cys His Pro Arg Leu Ser Leu His Arg Pro Ala Leu Glu Asp Leu
                565                 570                 575
Leu Leu Gly Ser Glu Ala Asn Leu Thr Cys Thr Leu Thr Gly Leu Arg
            580                 585                 590
Asp Ala Ser Gly Ala Thr Phe Thr Trp Thr Pro Ser Ser Gly Lys Ser
        595                 600                 605
Ala Val Gln Gly Pro Pro Glu Arg Asp Leu Cys Gly Cys Tyr Ser Val
    610                 615                 620
Ser Arg Val Leu Pro Gly Cys Ala Gln Pro Trp Asn His Gly Glu Thr
625                 630                 635                 640
Phe Thr Cys Thr Ala Ala His Pro Glu Leu Lys Thr Pro Leu Thr Ala
                645                 650                 655
Asn Ile Thr Lys Ser Gly Asn Thr Phe Arg Pro Glu Val His Leu Leu
            660                 665                 670
Pro Pro Pro Ser Glu Glu Leu Ala Leu Asn Glu Leu Val Thr Leu Thr
        675                 680                 685
Cys Leu Ala Arg Gly Phe Ser Pro Lys Asp Val Leu Val Arg Trp Leu
    690                 695                 700
Gln Gly Ser Gln Glu Leu Pro Arg Glu Lys Tyr Leu Thr Trp Ala Ser
705                 710                 715                 720
Arg Gln Glu Pro Ser Gln Gly Thr Thr Thr Tyr Ala Val Thr Ser Ile
                725                 730                 735
Leu Arg Val Ala Ala Glu Asp Trp Lys Lys Gly Glu Thr Phe Ser Cys
            740                 745                 750
Met Val Gly His Glu Ala Leu Pro Leu Ala Phe Thr Gln Lys Thr Ile
        755                 760                 765
Asp Arg Leu Ala Gly Lys Pro Thr His Ile Asn Val Ser Val Val Met
    770                 775                 780
Ala Glu Ala Asp Gly Thr Cys Tyr Arg Ser Glu Lys Asp Glu Leu
785                 790                 795
```

FIGURE 9 – 8